RICE UNIVERSITY

The Biophysical Basis for Adaptation: Predicting Evolutionary Outcomes from Physicochemical Properties

by

Andrés Salvador Benítez Cárdenas

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Yousif Shamoo
Professor of Biochemistry & Cell Biology,
Director of the Institute of Biosciences &
Bioengineering

Charles R. Stewart, Chair
Professor of Biochemistry & Cell Biology

Edward Nikonowicz
Associate Professor of Biochemistry &
Cell Biology

Jonathan Silberg
Associate Professor of Biochemistry &
Cell Biology

Laura Segatori
Assistant Professor of Chemical &
Biomolecular Engineering and
Biochemistry & Cell Biology

HOUSTON, TEXAS
November 2012

The Biophysical Basis for Adaptation: Predicting Evolutionary Outcomes from Physicochemical Properties

Andrés Salvador Benítez Cárdenas

November 28, 2012

Abstract

Experimental evolution can be used in conjunction with biophysical characterization of enzymes to determine the link between cellular fitness and physicochemical properties of enzymes. Sequencing of ancestral and evolved populations can be used to compare the outcomes of experimental evolution with measurements of fitness, using growth rate assays to correlate fitness outcomes to specific mutations. Combined with enzyme assays of kinetic properties that can provide a direct link between genotypic and phenotypic changes of adaptive mutants, we can model the complex relationship between genotypic changes and evolutionary outcomes.

Two experimental evolution systems were used to explore the link between enzyme properties and fitness outcomes. In the first series of studies, a "weak link" evolution experiment was used to explore the effect of reducing selection strength on altering accessible pathways for adaptation. In the "weak" link method the essential gene for adenylate kinase (AK) was replaced in the chromosome of the thermophile *Geobacillus stearothermophilus* with a homolog from *Bacillus subtilis*. Replacement with the maladapted gene confers a high fitness cost, and therefore mutations that restore function of AK are strongly favored. Two triple mutants of AK containing a new combination of single point mutants identified under strong selection, $AK_{Q199R/A193V/Q16L}$ and $AK_{Q199R/T179I/Q16L}$ were discovered through an adaptation experiment using a weak temperature ramp; suggesting that the adaptive landscape for AK thermostability is highly constrained. A thermostable coupled assay was developed for measuring adenylate kinase activity using LDH_{TT} and PK_{GST} at high temperatures. The triple

mutants had increased fitness compared with the double mutant ancestors, but the triple mutants displayed negative epistasis effects on fitness.

In the second experimental evolution system, a mathematical model was developed to investigate the role of adaptive mutations, in the tetracycline inactivation enzyme TetX2, on antibiotic resistance to minocycline. Growth rate measurements, enzyme kinetics, and flux balance analysis were used to develop a model to predict the effect on growth rates of TetX2 and seven adaptive TetX2 variants at different minocycline (MCN) concentrations. Population histogram measurements for the experimental evolution study were measured using a high throughput Illumina sequencing method(FREQ-SEQ). We found that the model was able to accurately predict the fitness outcomes for the wild type and the seven single mutants of TetX2 that were originally isolated, as well as for a double mutant that was not used in the development of the original model. The mathematical model accurately predicts that the two mutants $\mathrm{TetX2_{T280A}}$ and $\mathrm{TetX2_{N371I}}$ provide the largest fitness benefits, in agreement with the results of flask experiments on adaptation to MCN. The model was also able to accurately predict enzyme parameters from growth rate values, with a specific emphasis on predicting the ratio of $\mathrm{V_{max}/K_{M(MCN)}}$. The model allows us to make predictions about the fitness benefits of physicochemical changes to enzymes, and can be used as a high throughput method for determining enzyme kinetic parameters without requiring protein purification.

Understanding how physicochemical changes of enzymes relate to phenotypic changes, and ultimately to fitness, requires knowledge of both the molecular basis for determining enzyme properties, and how selection acts on fitness differences to determine evolutionary outcomes. This research provides direct links between physicochemical changes and adaptive phenotypes, as well providing observations of how

adaptive landscapes and fitness changes affect evolutionary outcomes.

Acknowledgments

I want to thank everyone who has made the work presented here possible. I'd like to start by thanking Dr. Yousif Shamoo for his support and guidance during the development of the thesis. He had the patience to allow me to find my own path, but without his knowledge about biochemistry, and his insights into evolution, a project with the scope and range of this thesis would not have been possible.

I would like to thank my committee, Dr. Steward, Dr. Nikonowicz, Dr. Silberg, and Dr. Segatori, for their input and suggestions, both on the thesis, and on my development as a researcher.

I want to thank members of the Shamoo laboratory, both past and present, for their help and advice in the work involved in this thesis. Thank you to Kasia Walkiewicz for her support and friendship, she has always been there when I needed to work out a problem, work related or not. She introduced me to the topic of TetX2 and minocycline resistance, and her efforts in identifying and characterizing the mutants were the foundation for the mathematical modeling. I want to thank Matt Peña for helping to guide me when I was starting out in the lab, especially with the work relating to adenylate kinase and working with the turbidostat, and for making the lab feel welcoming and inclusive for everyone. I want to thank Corwin Miller for his help and friendship. He is always willing to lend a hand when problems crop up in experiments, and generally makes the lab a more fun and interesting place to work at. I wish to thank Milya Davlieva for her help, especially with protein purification, and for always improving the mood of the lab with her cheerful attitude. I'd also like to thank Gerda Saxer for her advice and support. She introduced me to R and

helped to guide me through learning it, and is always willing to listen to any concerns I have and give useful advice. I want to thank Anisha Perez for her friendship and help with the TetX2 project, and I hope the work presented here is able to help her with her own research on TetX2, or wherever her project leads her. Thank you to Troy Hammerstrom, Kathyrn Beabout, and Xu Wang for their help and support as well. I also want to thank all of the undergraduate students that have worked in the Shamoo lab over the years I've been there. Without them the lab would be a lot more quiet and a lot less interesting. I especially want to thank Colin Bacorn, Christina de Bruyn Kops, Wanna Zhang, Rebecca Jeun, and Christine Sun who were all directly involved in helping the project at one point or another.

I would also like to thank Aaron Collier and Andria Denmon for their friendship and support, and for the random conversations we had in the hallways of Keck. I would like to thank the other labs that we share the building with, the Tao lab, the Nikonowicz lab, the Olson lab, the Bennett lab, and the Silberg lab, for both their help with experimental questions and laboratory supplies, and for making Keck Hall an enjoyable place to work at.

I am very grateful for the love and the support that I have received from my family and friends during the last few years, both from those nearby and those far away. I want to thank my parents, Salvador and Rosy, for their unconditional support for my education, intellectually and emotionally. They are always confident in me and my ability, even when I am not. I also want to thank my brothers, Mauricio and Diego, and my sister Regina, for their love and their concern. Even when they are not able to completely follow my discussions about my research they are always interested in listening to what I am doing, and willing to giving me advice on work and life.

Thank you again to everyone who has helped me, I am deeply grateful for your

support. The writing of the thesis is an individual effort, but writing it allowed me to remember and appreciate everyone's contributions that have made it possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Mutations of Small Effect Drive Adaptation

Adaptive evolution is driven by the accumulation of changes at the molecular level that affect the phenotype of the organism. Mutations that increase enzyme efficiency and cause incremental fitness gains are enriched through natural selection. The understanding of the importance of small changes at the molecular level predates the discovery of genetics. Darwin recognized that nature cannot make spontaneous jumps in evolution; natural selection must act on small variations between organisms [Darwin, 1859]. However, he was unable to provide a mechanism for how infinitesimal changes in variation become fixed in a population.

Kimura proposed that for a mutation to become fixed in the population it must not only provide a beneficial effect, but the fitness gain has to be large enough to prevent the mutation from being lost through random effects [Kimura, 1968, Orr, 2005]. The probability of a mutation with a beneficial effect being lost by random drift is dependent on selection strength and mutation frequency. Random genetic drift

and not adaptive evolution becomes a dominant factor in determining gene evolution if fitness benefits are not strong enough to overcome random loss of adaptive mutants. Alternatively, mutations that cause such a small change in protein function so as have effectively a neutral fitness effect, can only become fixed in the population through random chance.

Fisher was the first to propose a mechanism to link variation at the genetic level to variation at the phenotypic level [Fisher, 1930]. Fisher proposed that beneficial mutations of small phenotypic effect are more frequent than beneficial mutations of larger phenotypic effect (1.1). Because mutations of larger phenotypic effect have higher probabilities of disrupting protein function, adaptation would be dominated by mutations of small phenotypic effect [Orr, 2006]. Since mutations that cause a small change in phenotype are more numerous that mutations of greater phenotypic effect, mutations with small effects on fitness will accumulate faster than large beneficial mutations under non-selective conditions. Mutations that have a higher effect on fitness, although rare, have a higher probability of becoming fixed because they are less likely to be lost under selection [Kimura, 1968]. Fisher's geometric model of the distribution of beneficial mutations provided a foundation for understanding the mechanisms of evolution.

1.2 Mutation Rates Drive Adaptation

Under constant environmental conditions most genes under selection exist in a state that provides high fitness, and because mutations will be predominately comprised of single nucleotide mutations, adaptation to small disturbances in environment will mostly be concerned with the extreme end of Fisher's geometric distribution of mu-

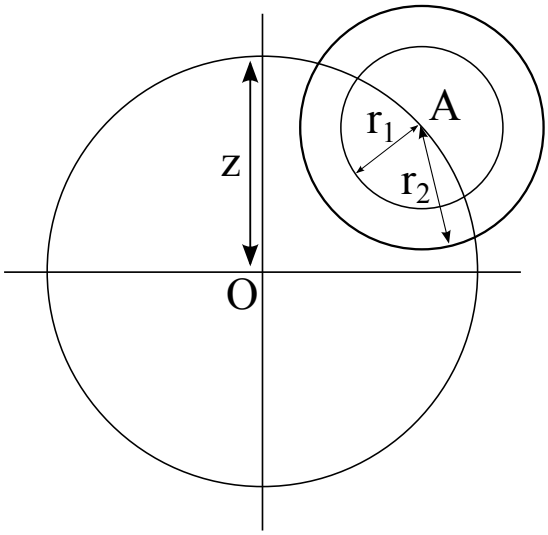


Figure 1.1: In this example an organism with two degrees of phenotypic change sits at point A that is a distance z from a fitness optimum at point O. Mutations of phenotypic effect of size r will move the fitness of the organism randomly. Smaller phenotypic effect mutations (r1) have a higher probability of moving towards the fitness optimum than mutations of larger phenotypic effect (r2). Only mutations that move the fitness of the organism inside the larger circle of radius z will be beneficial. As the radius of circle r become smaller, corresponding to mutations of smaller size, the fraction of mutations that will have beneficial effects will be larger [Fisher, 1930].

tations [Orr, 2006]. The mutation rate of *Escherichia coli* is about $8.9 * 10^{-11}$ mutations per base per generation for single nucleotide substitutions [Wielgoss et al., 2011]. This translates to a mutation rate of $4 * 10^{-5}$ mutations per *E. coli* genome per generation. The mutation rate for spontaneous double mutants is much smaller at $4 * 10^{-5} \times 4 * 10^{-5}$. Insertions and deletions tend to be rare except at sequence repeats, where they appear at a rate that is 1/10th the rate of single base pair substitutions [Lee et al., 2012]. Therefore, except for rare occurrences, natural selection will mostly be driven by single substitutions [Orr, 2005, Gillespie, 1986].

Experiments involving mutator strains have shown that under stressful conditions there is a linear increase in adaptation rate in response to an increase in mutation rate [Arjan G et al., 1999]. If the population size is small enough that the rate of adaptation is limited by the availability of beneficial mutations,f mutator strains will increase the adaptation rate of a population [Gonzalez et al., 2008]. Mutator phenotypes can then become enriched in the population by piggy-backing with adaptive mutations [Denamur and Matic, 2006]. If the endogenous mutational diversity is high enough, then there will be a sufficient amount of beneficial mutations present in the population and the adaptation rate will not increase further [Arjan G et al., 1999]. Clonal interference, where multiple beneficial mutations compete in the same population, prevents further increase in adaptive rates since only the mutation providing the largest fitness increase will rise to fixation [Tanaka et al., 2003].

Mutation rate has been shown to increase under stress, leading to an increase in rates of adaptation [Galhardo et al., 2007, Gonzalez et al., 2008, Denamur and Matic, 2006]. Mutator phenotypes are not beneficial under stable environments because of the possibility for the accumulation of detrimental mutations [Chattopadhyay et al., 2009]. However, under stressful conditions, it can become beneficial to increase the

mutation rate to access rare beneficial mutations [Galhardo et al., 2007]. Rosenberg has shown that under environmental stress the mutation rate of nucleotides near an induced double stranded break on the Lac^+ gene increased up to 4 fold [Gonzalez et al., 2008]. Under conditions of stress the fidelity of the double strand break repair mechanism is reduced causing a local increase in mutation rates. Other studies have shown there are multiple mechanisms to increase mutation rates in response to environmental stress [Galhardo et al., 2007].

1.3 Fitness Gains Taper After Initial Adaptive Steps

When an organism is introduced to a new environmental stress mutations of large phenotypic effect are favored because there is a large fitness gap between the original gene sequence and potential mutational outcomes. Once beneficial mutations start to accumulate the speed of evolution tapers as fitness gains become smaller [Elena and Lenski, 2003, Chou et al., 2011, Khan et al., 2011, Nagel et al., 2012]. Due to neutral drift, small mutations continue to accumulate at a linear rate although the fitness of the population plateaus. [Elena and Lenski, 2003, Yu et al., 2009].

In a long-term evolution experiment by Richard Lenksi, replicate populations of *E. coli* were adapted to grow in minimal media lacking glucose but containing citrate as a carbon source [Elena and Lenski, 2003]. *E. coli* cannot utilize citrate as a carbon source, so a gradual increase in fitness was observed except for one of the replicate populations where the bacteria were able to evolve the ability to utilize citrate. A fitness jump was observed when the bacteria accumulated the required mutations that allowed the evolution of citrate utilization. The accumulation of potentiating mutations was necessary to allow access to adaptive pathways that provided significant

increases in fitness. [Blount et al., 2008, Rendel, 2010]

1.4 Selection Strength Affects Accessibility of Adaptive Mutations

Wright proposed the idea of an adaptive landscape as a method to visualize the surface on which evolution happens [Wright, 1988]. A different version of the adaptive landscape was proposed by Smith, where, because of limitations set by the mutation rate, proteins would typically adapt by single base substitutions [Smith, 1970]. While Smith's model was concerned with protein sequences, his concept can be also be applied to adaptation of DNA sequences [Orr, 2009, Sasaki and Nowak, 2003]. The adaptive landscape would then be composed of all possible genotypes so that neighboring sequences differ by single gene replacements [Orr, 2009]. At each step of adaptation, natural selection would be able to sample only the sequences that differ by a single mutation [Orr, 2006]. Degeneracy in codon usage further limit the amino acid replacements accessible from single mutations [Kimura, 1968]. Fitness can be projected onto the surface such that the relative height of each node represents fitness differences [Cooper and Lenski, 2010]. Natural selection will drive the organism up the nearest fitness peak through the accumulation of beneficial mutations [Orr, 2005].

The shape of the adaptive landscape is also affected by interactions between individual mutations [Perfeito et al., 2011, Chou et al., 2011, Lunzer et al., 2010]. Since most mutations have epistatic interactions the adaptive landscape will tend to be rugged, containing multiple local fitness peaks [Wright, 1988]. Epistasis is defined as the non-additive effects of two individual mutations on phenotype [Elena and Lenski, 2003]. The phenotypic effect from two mutations can be greater or lesser than the

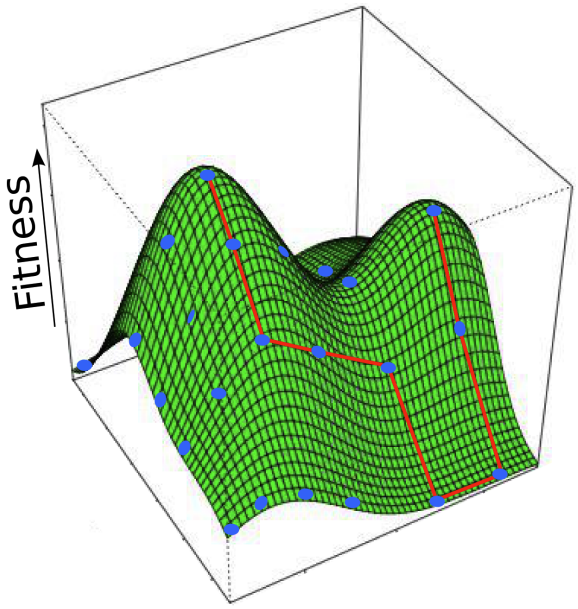


Figure 1.2: Adaptive landscape with two fitness peaks. Each node represents a single mutational difference, pathways are limited to horizontal or vertical movements. A two dimensional landscape represents a very simplified system with variation limited two possible mutations at each step. Two different pathways are shown, each leading to a local fitness optimum. Once a fitness peak is reached, movement across the landscape is limited since mutations will be detrimental.

additive effect of the individual mutations because of interactions between the two mutations. Beneficial mutations show negative epistasis as they accumulate leading to a decrease in adaptation rate [Chou et al., 2011, Nagel et al., 2012]. Additive effects on enzyme function decrease in fitness benefit as benefits from improving individual phenotypic effects become saturated [Khan et al., 2011]. As the organism approaches a fitness peak the beneficial effect of individual mutations becomes smaller due to epistasis [Hartl et al., 1985]. However, mutations that potentiate beneficial mutations can appear through positive epistasis [Lunzer et al., 2010].

Positive epistasis between inter-protein mutations is commonly observed when background mutations are required to allow a secondary mutation to have a beneficial effect on fitness [Weinreich et al., 2006, Kryazhimskiy et al., 2009]. For example, compensatory mutations can increase the stability of a protein thus allowing for mutations that increase activity but can have a destabilizing effect on the protein [Couñago et al., 2008, Miller et al., 2006, Tokuriki et al., 2008]. The combined beneficial effect of both mutations is higher since the stabilizing mutation has a neutral effect on fitness by itself, and the initial mutations would have a decreased fitness benefit without the compensatory mutation [Levin et al., 2000]. Epistasis has a role in limiting available pathways for evolution since the mutational background affects the benefits of subsequent inter-protein and intra-protein mutations [Weinreich et al., 2006].

Since the protein must remain functional at each step during adaptation, only sequences that provide neutral or beneficial effects can become fixed in the population [Orr, 2005]. The adaptive walk of a protein sequence can be described as the sequential progress of a protein from an area of low fitness to a local fitness peak due to natural selection [Smith, 1970]. Selection strength, reflected by the slope of the fitness peak, will affect allowable intermediates for each step in the pathway [Eyre-

Walker and Keightley, 2007]. Under a changing environment each intermediate must be functional since only steps that are able to keep up with environmental changes will be allowed [Collins et al., 2007]. Reducing the selection strength by limiting the rate of environmental change allows mutations that provide smaller fitness benefits, or neutral fitness effects, to persist in the population. Adaptive pathways that stem from initial weak mutations will be accessible under weaker selection conditions. The height of fitness peaks and the ruggedness of the landscape will therefore be affected by selection strength.

1.5 Protein Function Balances Activity and Stability Trade-Offs

There is an inherent trade off in activity and stability for protein function. Mutations in the active site of a protein that increase activity have a tendency to be destabilizing [Tokuriki et al., 2008]. Generally mutations that provide new function, such as changing enzyme specificity, consequently have decreased stability [Wang et al., 2002]. Increasing the rigidity of a protein as a result of an increase in stability, on the other hand, can have an adverse effect on activity [Arnold et al., 2001]. Experiments using thermostable proteins have shown that an increase in stability often causes a corresponding shift in the activity maximum of the enzyme [Couñago et al., 2008]. Conversely, psychrophilic adaptation has been shown to lead to increased flexibility of enzymes [Roovers et al., 2001]. Increasing stability has also been shown to enhance the ability of enzymes to tolerate adaptive mutations [Bloom et al., 2006,Draghi et al., 2010]. Robustness to destabilization from mutation has been shown to be a feature of natural enzymes [Taverna and Goldstein, 2002]. Specifically increased protein stabil-

ity can buffer proteins from the deleterious effects of destabilizing mutations [Bloom et al., 2005]. Because most mutations are destabilizing under neutral mutation stability will be maintained by the accumulation of compensatory mutations in response to destabilizing mutations accumulated through drift [Serrano et al., 1993, Soskine and Tawfik, 2010]. Proteins under neutral evolution exist under equilibrium between the necessity of flexibility for enzyme function and the requirements of excess stability to provide a buffer to the accumulation of destabilizing mutations [DePristo et al., 2005]. Adaptation to novel substrates [Wang et al., 2002] and adaptation to changing environments [Couñago et al., 2008] both often require mutations that have trade-offs between activity and stability for adaptation.

1.6 Historical Contingency Affects the Stochasticity of Evolution

The existence of multiple pathways for adaptation highlights an interesting question about the repeatability of evolution [Gould and Eldredge, 1977]. How constrained are outcomes of evolution if multiple pathways exist? Lenski's 12 experimental populations [Lenski et al., 1991] provide useful insights into the role of stochasticity in determining evolutionary outcomes. They determined that the 12 experimental populations shared mutations in similar genes or pathways, as well as having similar increases in fitness [Yu et al., 2009, Lenski et al., 1991]. Other experiments have also shown that under strong selection there are a limited amount of pathways available for evolution [Wichman et al., 2000, Wood et al., 2005]. However, increasing the environmental heterogeneity of the adaptive landscape produces a more rugged landscape and, as a consequence, evolutionary outcomes become more stochastic [Cooper and

Lenski, 2010].

Lenski was also able to provide proof for the role of historical contingency in evolutionary outcomes [Blount et al., 2008] . If evolutionary outcomes are constrained the starting point of a population becomes important in determining possible outcomes of evolution [Travisano et al., 1995]. They were able to show that mutations allowing citrate metabolism were only beneficial in certain backgrounds of evolved populations [Yu et al., 2009]. Other experiments have also shown that mutations with neutral or small effects can be useful in allowing access to novel pathways for evolution [Rendel, 2010].

1.7 A Biophysical Basis for Adaptation

The causal relationship between physical properties of enzymes and their relative effects on fitness must be established to fully understand the rules that govern natural selection. The interaction between genotype and phenotype is complex and causality is hard to determine [Dykhuizen et al., 1987]. Different genotypes can have the same phenotypic effect, and the same genotype under different environments can have different effects on phenotype [Dykhuizen and Dean, 1990]. Determination of functional polymorphisms is further complicated by the large fitness effect that can occur from relatively small functional effects [Eanes, 1987]. However, the system can be simplified by controlling experimental conditions in laboratory evolution experiments [Kawecki et al., 2012, Couñago et al., 2006]. Evolution experiments can be designed to target specific phenotypic changes by controlling selection conditions [Lenski et al., 1991]. Population size [Campos and Wahl, 2009, Barrett et al., 2006] and mutation rate [Camps et al., 2003, Gonzalez et al., 2008, Chattopadhyay

et al., 2009] can also be used to control mutation frequency as well. Enzyme variants from experimental or natural populations can be characterized to measure physical properties such as protein stability and activity [Couñago et al., 2006]. Differences in variation of enzyme properties can then be compared to population measurements to make correlations about the fitness effects of enzyme variants [Peña et al., 2011].

The mechanistic approach to studying evolution was described by Dean and Thornton as the "functional synthesis" [Dean and Thornton, 2007]. Dean and Thornton described a growing synergy between approaches in molecular biology, biochemistry , and experimental evolution to elucidate the complex relationship between genotype and phenotype. Hypothesis generated from the results of evolution experiments can be tested through biophysical comparisons of enzyme variants and their resulting effects on phenotype [Storz and Wheat, 2010] . The development of high throughput DNA sequencing allows us to identify changes in gene sequences as well as provides mechanisms for measuring population distributions [Ragoussis, 2006, Chubiz et al., 2012]. Genes from evolved populations can be cloned into the ancestral genome, isolating any change in fitness from the targeted gene [Stepanova et al., 2008, Couñago et al., 2006, Weinreich et al., 2006]. High resolution protein structures determined by X-ray crystallography or NMR can be coupled with measurements of enzyme kinetics to determine the physicochemical properties of enzymes [Couñago et al., 2008, Peña et al., 2010, Newcomb et al., 1997].

1.8 Flux Analysis Can Predict Fitness Outcomes

Mathematical models of metabolite flux through the cell can be used to predict growth rates and fitness [Lewis et al., 2010, Peña et al., 2011]. Measurable changes in the

physical properties of cells are correlated to measurable changes in overall fitness through a growth rate function based on metabolic flux analysis [Dykhuizen and Dean, 1990, Orth et al., 2010]. The flux through a metabolic pathway is related to the activity of the individual enzymes [Dykhuizen et al., 1987]. The fitness function for flux through a limiting metabolic pathway will be concave as the effects of increasing flux thorough an enzyme will diminish as metabolism becomes saturated [Hartl et al., 1985]. Dykhuizen was able to measure the relative effects on fitness of beneficial mutations in a lactose utilization pathway by comparing lactose utilization mutants isolated from adapting *E. coli* to grow under lactose limited media in a chemostat. He was able to show that an increase in β-galactosidase permease activity had a larger effect on fitness than the same increase in activity in β-galactosidase [Dykhuizen et al., 1987]. The rate of flux through multiple metabolic systems can also be shown to be proportional to the relative amounts of enzyme concentrations [Rossell et al., 2011], allowing analysis of effects of phenotypic variation on the genotype scale [Lerman et al., 2012]. Analysis of metabolite flux provides a powerful system for understanding correlations between the physical properties of enzymes and behaviors of populations undergoing evolution [Eanes, 1987, Dekel and Alon, 2005, Lunzer et al., 2002, Ibarra et al., 2002].

1.9 Focus of the Work

The goal of this dissertation is to unite physical changes to fitness and, in turn, to predict success in an evolving population. We are interested in following the path from molecular changes to their effect on the overall fitness of the organism. To understand the causal pathway from protein changes to fitness changes we must be able

to compare fitness measurements to measurable changes in physical properties of enzymes. Understanding the complete pathway relies in uniting "top down" approaches that elucidate the end results of adaptation, with "bottom up" approaches that seek to provide a reasoning for the mechanisms that provide fitness differences. Random mutagenesis and experimental evolution techniques were used to provide a source of enzyme variants with measurable fitness differences. Enzyme kinetics and stability measurements were then used to uncover physical properties of enzyme variants to determine the link between physical properties and fitness. We compared the properties of enzymes to measurements of population dynamics during evolution to provide insights into mechanisms determining evolutionary outcomes.

We elucidated the link between evolutionary outcomes and enzyme properties using two different experimental evolution systems. Evolution of thermostability in a turbidostat under weak selection conditions was used to test how the ruggedness of the adaptive landscape affects possible outcomes for evolution. Measurements of the shift of the temperature profile for stability and activity in response to changing temperature were then used to investigate the effect of the ruggedness of the fitness landscape on determining adaptive walks. Mathematical modeling was then used to investigate the role of small changes in physical properties of enzymes on determining fitness changes. We used the mathematical model to determine a fitness function to predict growth rates for an enzyme that provides resistance to growth rate inhibition by antibiotics and used enzyme variants to determine model parameters. We were then able to show that the same model can be used to determine kinetic properties from growth rates.

1.10 The "Weak Link" Method for Studying Evolution

Replacement of an essential gene in the chromosome with a maladapted homologue can focus adaptation of the targeted gene under the appropiate selection conditions. The replaced gene becomes a "weak link" preventing the cell from growing at maximum growth rate [Couñago et al., 2006]. Mutations that restore function of the gene will provide a large fitness benefit and will be strongly favored. The essential protein adenylate kinase was replaced in *G. stearothermophilus* with the homologue from *B. subtilis* to create our "weak link" system [Couñago et al., 2006]. Adenylate kinase was chosen as our target enzyme because it is necessary for maintaining adenylate homoestasis [Atkinson, 1968]. Growth rate can be shown to be dependent on the concentration of high energy phosphorylated adenylate species present in the cell, which is dependent on proper adenylate kinase function [Swedes et al., 1975]. The energy charge of the cell (EC) is defined as the fraction of high energy phosphorylated adenylate species to lower energy species [Atkinson, 1968], is the availability of the high energy phospohrylated species [Atkinson and Walton, 1967].

$$\mathrm{EC} = \frac{\mathrm{ATP} + \frac{1}{2}\mathrm{ADP}}{\mathrm{ATP} + \mathrm{ADP} + \mathrm{AMP}} \tag{1.1}$$

Adenylate kinase is required to regenerate ADP by catalyzing reaction R 1.1.

$$\mathrm{ATP} + \mathrm{AMP} \rightleftharpoons 2\,\mathrm{ADP} \tag{R 1.1}$$

ATP can then be regenerated by ATP synthases.

$$\mathrm{ADP} + \mathrm{P_i} \rightleftharpoons \mathrm{ATP} + \mathrm{H_2O} \qquad \text{(R 1.2)}$$

Without proper adenylate kinase function the cell lacks other efficient mechanisms for regenerating ADP, leading to an inability to produce ATP and a decrease in the EC of the cell [Peña et al., 2010]. The function of adenylate kinase is tightly controlled since adenylate nucleotide concentrations are essential regulators of cellular metabolism [Hardie and Hawley, 2001, Atkinson and Walton, 1967]. Proper adenylate kinase function is therefore essential for cell survival, and furthermore, a decrease in AK function will lead to a proportional decrease in cell metabolism because of the reduction in EC of the cell [Atkinson, 1968, Swedes et al., 1975, Peña et al., 2010].

The crystal structure of adenylate kinase from *B. subtilis* reveals that the catalytic domain consists of two binding pockets for the two substrates and a lid domain [Saint Girons et al., 1987]. Flexibility of the enzyme, and especially of the lid domain, has been shown to be essential for proper function of the enzyme [Vonrhein et al., 1995]. Adenylate kinase adaptation to function at high temperatures requires a balance of increased thermal stability without compromising the required flexibility of the enzyme [Couñago et al., 2008]. The lid domain closes over the site of phosphoryl transfer to prevent hydrolysis of ATP by water [Müller et al., 1996].

A temperature sensitive strain of *G. stearothermophilus* NUB3612-R was created through a chromosomal replacement of $\mathrm{AK_{(GSTE)}}$ with $\mathrm{AK_{(BSUB)}}$ [Couñago and Shamoo, 2005]. Comparison of the structures of *G. stearothermophilus* AK ($\mathrm{AK_{(GSTE)}}$) and *B. subtilis* AK ($\mathrm{AK_{(BSUB)}}$) reveals that the thermophilic enzyme is more rigid at lower temperatures due to an increase in the strength of electrostatic interactions [Bae and Phillips Jr, 2004]. The two enzymes share 74% amino acid sequence and have

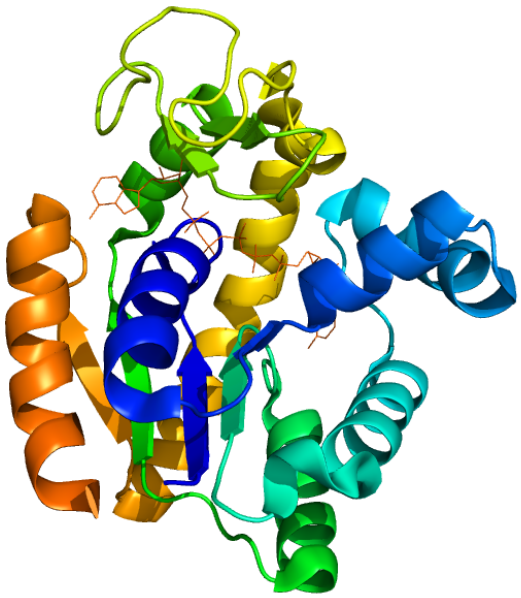


Figure 1.3: Structure of adenylate kinase mutant Q199R in complex with the inhibitor diadenoside pentaphosphate (Ap5A) [Couñago et al., 2006]. Ap5A binds to both the ATP and AMP binding pockets. The structure is presented in the closed conformational state [Saint Girons et al., 1987].

similar structural identity, however, the thermal denaturation midpoints (T_M) are significantly different: $AK_{(BSUB)}$ T_M is 47.6℃ and $AK_{(GSTE)}$ T_M is 74.5℃ [Bae and Phillips Jr, 2004]. The growth range for strain *G. stearothermophilus* NUB3621-R:ThEV is limited to below 55 ℃ compared to the wild type optimal growth temperature of 68℃ [Couñago and Shamoo, 2005]. Mutants that rescued the temperature sensitive phenotype of *G. stearothermophilus* NUB3621-R:ThEV were recovered through an evolution experiment utilizing a turbidostat. A turbidostat is a type of bacterial continuous batch reactor where the cell density is kept constant by regulating the nutrient flow into the reactor vessel [D E Dykhuizen, 1983]. The bacteria were grown under a temperature ramp of 0.5℃ per day from an initial temperature of 55℃ over 30 days to a final temperature of 70℃ [Couñago et al., 2006]. The mutant AK_{Q199R} was isolated at 55℃ from the first sample taken from the reactor vessel. $AK_{(BSUB)Q199R}$ has significantly increased stability and activity compared to WT $AK_{(BSUB)}$ at 55℃. $AK_{(BSUB)}$ has a maximum activity that is between 55℃ and 60℃ [Peña et al., 2011, Couñago et al., 2008]. The population in the vessel remained stable until 60℃ when $AK_{(BSUB)Q199R}$ was replaced in the population by a series of double mutants. The double mutants competed in the population with varying degrees of success until 68℃ when one double mutant, $AK_{(BSUB)Q199R/Q16L}$, became fixed in the population. The triple mutant increased the stability over 10℃ compared with WT $AK_{(BSUB)}$ [Couñago et al., 2006]. Further studies of the 6 $AK_{(BSUB)}$ variants show that each of the mutants shifts the temperature optimum for each mutant to a higher temperature, with $AK_{(BSUB)Q199R/Q16L}$ showing the largest shift in activity [Peña et al., 2011].

1.11 Antibiotic Resistance Provides a Controlled Environment to Study Adaptation

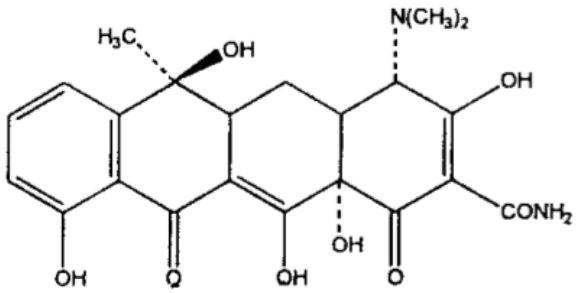
Antibiotic resistance is a compelling target for laboratory evolution experiments both because it provides clinical relevance, and selection strength can be changed readily. [Elena and Lenski, 2003, Dykhuizen and Dean, 1990, Charusanti et al., 2012]. The rise of antibiotic resistance over the last 50 years due to widespread use [Livermore, 2003] has led to increased interest in the discovery of novel antibiotics and antibiotic targets [Cohen, 2000, Boucher et al., 2009]. There is also a resurgence of emphasis on the control and treatment of infectious diseases due to the accelerating spread of multi-drug resistant bacteria and the reduction in the development of novel therapeutic antibiotics [Cohen, 2000, Boucher et al., 2009]. The global rise during the last decade of multi drug resistant *Staphylococcus aureus* [Boucher et al., 2009], drug resistant respiratory pathogens [File, 2006], and multi and pan resistant Gram-negative bacilli [Giske et al., 2008] underscores the importance of antibiotic resistance at the forefront of health care issues [Spellberg et al., 2008]. The role of broad spectrum antibiotic resistance enzymes and membrane efflux pumps in providing multi-drug resistance to bacteria highlights the importance of focusing on understanding mechanisms for antibiotic resistance [Alekshun and Levy, 2007]. Laboratory evolution experiments provide a unique environment to explore both the mechanisms bacteria use to adapt to antibiotics, as well as to help predict new sources of antibiotic resistance. Experiments using directed evolution have been able to recapitulate the appearance of mutations that have clinical relevance [Weinreich et al., 2006, Orencia et al., 2001] as well as to help understand the interactions between mutations that confer resistance [Oelschlaeger, 2008].

Development of antibiotic resistance provides a good system for investigating mechanisms of adaptation such as, the role of compensatory mutations in reducing the cost of antibiotic resistance [Lenski, 1998, Wang et al., 2002], or the role of mutation strains on enhancing rates of adaptation [Levin et al., 2000, Bjorkholm et al., 2001]. Resistance to antibiotics can be used to provide a strong selection to allow the isolation of adaptive mutants [Barlow and Hall, 2002, Orencia et al., 2001]. Selection can also be controlled by changing antibiotic concentration or changing the rate of the antibiotic concentration ramp. Fitness gains can be easily quantified by measuring growth rate under different antibiotic concentrations [Weinreich et al., 2006, Vakulenko et al., 1998]. Stemmer showed that experimental evolution using β-lactamase can be used to predict possible pathways for antibiotic resistance. He used DNA shuffling and recombination to find variants of TEM-1 β-lactamase with increased resistance to cefotaxime [Stemmer, 1994]. He found that the triple mutant E104K/M182T/G238S produced a 500-fold increase resistance over the wild type sequence. This same triple mutant was identified as clinical isolate TEM-52 [Poyart et al., 1998]. The available pathways for antibiotic resistance seem to be highly constrained, comparable to other natural and directed evolution experiments, presumably due to strong selection conditions [Wichman et al., 2000]. Weinreich et al., used a library of combinations of these and other well known TEM-1 mutations to investigate the role of epistasis in antibiotic resistance [Weinreich et al., 2006]. They created all possible combinations of 5 mutations to the TEM-1 β-lactamase gene, and measured the minimum inhibiting concentration (MIC) of cefotaxime for each combination as a proxy for fitness. They were able to show that epistasis between the 5 mutations limited available pathways for adaptation to less than 10% of the total available outcomes. Evolution of antibiotic resistance has also been used to answer

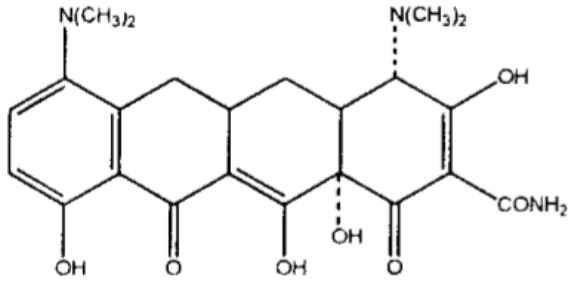
questions about the mechanisms that drive adaptation [Barlow and Hall, 2002, Zhang et al., 2011]. Our goal is to use adaptation to the tetracycline derivative minocycline by bacteria containing the tetracycline resistance protein TetX2 to create a growth rate function that can predict how functional differences in protein function translate to antibiotic resistance.

1.12 Modeling of Mechanisms of Evolution using Antibiotic Resistance

Tetracycline class antibiotics are broad spectrum bacteriostatic antibiotics affecting both Gram positive and Gram negative bacteria [Chopra and Roberts, 2001]. Tetracycline antibiotics inhibit protein synthesis by binding to the A (aminoacyl) site of the 30S ribosomal protein blocking binding of amino-acyl tRNA [Brodersen et al., 2000, Chopra and Shales, 1981, Spahn and Prescott, 1996]. Specifically tetracycline binds to the sugar phosphate backbone of the 16S RNA [Brodersen et al., 2000]. There are, however, other secondary sites where tetracycline is known to interact with the 30S subunit, and it is unknown if these sites are relevant for antibiotic activity [Brodersen et al., 2000, Pioletti et al., 2001]. Tetracyclines cross the outer membrane of gram negative bacteria through the OmpF or OmpC porin channels as a positively charged complex with magnesium (Mg^{2+}) [Nikaido and Thanassi, 1993]. The Donner potential, the unequal distribution of ionic species across a membrane, causes an accumulation of tetracycline in the periplasm [Thanassi et al., 1995]. Tetracyclines dissociate from Mg^{2+} once in the periplasm [Thanassi et al., 1995]. Since tetracycline is weakly lipophilic the uncharged species can cross the inner membrane into the cytoplasm [Sigler et al., 2000]. Tetracycline becomes chelated to Mg^{2+} in



(a) Tetracycline

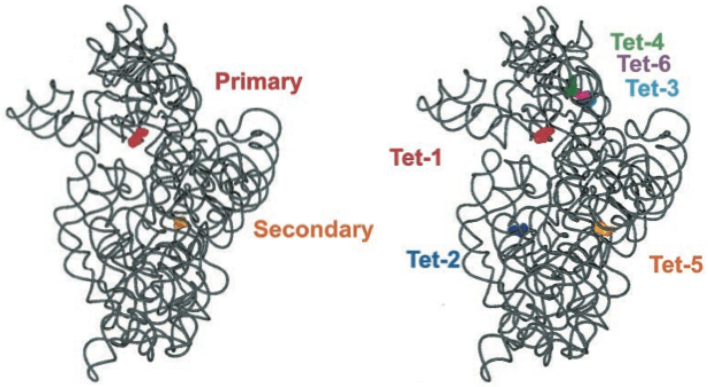


(b) Minocycline

Figure 1.4: The structures of a) tetracycline and b) minocycline. The hydrophilic functional groups and the 4 ring system are conserved among tetracycline variants and required for function. Reprinted by permission from the American Society for Microbiology: Microbiology and Molecular Biology Reviews [Chopra and Roberts, 2001] Copyright (2001).

the cytoplasm due to the pH gradient across the inner membrane, causing further accumulation inside the cytoplasm [Chopra and Roberts, 2001, Nikaido and Thanassi, 1993, Sigler et al., 2000]. Minocycline is a tetracycline derivative that has increased hydrophobicity compared with tetracycline [Chopra and Roberts, 2001], resulting in increased transport into the cell .

Resistance mechanisms for tetracycline can be separated into groups by their mode of action. Most of the genes conferring tetracycline resistance are in the form of efflux pumps, which protect the ribosome by reducing the intracellular concentration of tetracycline, and including TetA. Ribosomal protection proteins, such as TetM, work by preventing binding of tetracycline to the ribosome either by directly inhibiting



(a) (b)

Figure 1.5: Tetracycine shows either a) 2 [Brodersen et al., 2000] or b) 6 [Pioletti et al., 2001] binding sites for the 30S ribosomal subunit. The primary site in figure a and the Tet-1 site in figure b show nearly identical binding interactions, while the secondary site and Tet-5 site show similar but distinct binding [Pioletti et al., 2001]. The primary site is near the amino-acyl tRNA binding site (site A) of the ribosome and is predicted to be the main source of ribosome inhibition [Brodersen et al., 2000]. Reprinted by permission from American Society of Microbiology: Antimicrobial Agents and Chemotherapy [Connell et al., 2003], copyright (2003).

tetracycline binding or by promoting the release of tetracycline after binding to the ribosome [Connell et al., 2003]. TetX2 is the only known tetracycline resistance protein that works by enzymatic degradation of the antibiotic [Chopra and Roberts, 2001].

TetX was first identified in two related transposons, Tn*4351* and Tn*4400*, from *Bacteroides fragilis* [Speer et al., 1991]. Two other *tetX* homologues (*tetX1* and *tetX2*) were identified in the transposon CTnDOT from *Bacteroides thetaiotaomicron* [Whittle et al., 2001]. The first of the two *tetX* genes in CTnDOT, *tetX1* shares 66% amino acid identity to TetX, however, *tetX2* shares 99% amino acid identity with TetX [Speer et al., 1991]. TetX2 regioselectively hydroxylates tetracycline at the C11a position [Yang et al., 2004]. The enzyme is a monooxygenase-containing

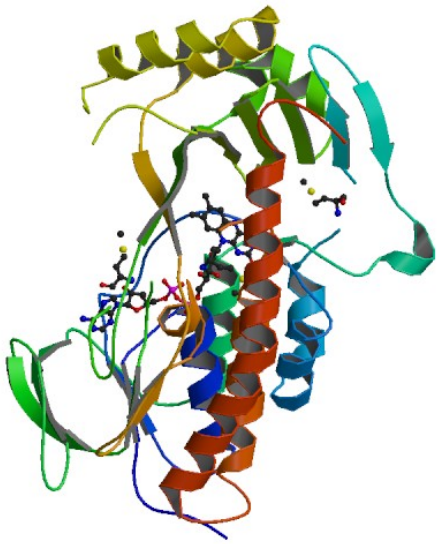


Figure 1.6: Structure of TetX2 and FAD cofactor with substrate analog [Walkiewicz et al., 2012].

a flavin cofactor, which is required for proper function [Moore et al., 2005]. Upon hydroxylation, tetracycline is unstable and degrades non-enzymatically [Yang et al., 2004]. The enzyme requires NADPH and O_2 to re-oxidize the flavin cofactor after each reaction [Yang et al., 2004]. The tetracycline derivative minocycline is a good target for adaptation since wild type *tetX2* has low activity towards minocycline compared with other tetracyclines [McMurry et al., 1982].

1.13 Goals for the thesis

We used the "weak link" method to investigate the effects of selection strength on accessible adaptive pathways. Using the thermolabile strain of *Geobacillus stearothermophilus*, we isolated novel mutants of *adk* by modifying the selection strength of the turbidostat experiments. We also developed a novel assay for measuring enzyme ki-

netics of the thermostable adenylate kinases. This assay was used to measure the activity of the novel mutants of adenylate kinase and, coupled with measurements of stability, used to compare the properties of the enzymes identified from the weak selection experiment with those identified previously.

The second goal of the thesis was to develop a growth rate function to predict the effects of changes in the enzymatic properties of TetX2 on degradation of minocycline on and consequently, on growth rates and fitness. Growth rates and kinetic measurements of 8 mutants that were identified though directed evolution were used to develop a growth rate function to predict fitness effects. Enzyme properties of mutants were predicted from growth rates using the growth rate function to prove the reversibility of the model. The relative importance of $K_{M(MCN)}$. $K_{M(NADPH)}$, and V_{max}on determining growth rates was determined by analyzing the response of the model to changes in these parameters.

Chapter 2

Materials and Methods

2.1 Construction of plasmids for the expression of *Thermus thermophilus* L-lactate dehydrogenase (LDH_{TT}) and *Geobacillus stearothermophilus* pyruvate kinase (PK_{GST})

T. thermophilus genomic DNA was a gift from Dr. J. Silberg (Rice University). The lactate dehydrogenase gene (GI:55772495) was cloned into pET-28b (Novagen) with an N terminal HIS-tag ($pLDH_{TTHERMOPHILUS}$). The pyruvate kinase gene from *Geobacillus stearothermophilus* (GI:285623) was synthesized (Genescript) after codon optimization for expression in *Escherichia coli*. The synthetic gene was subcloned into pET-28b with an N terminal HIS-tag for expression in *E. coli* using NdeI and HindIII restriction sites ($pPK_{GSTEAROTHERMOPHILUS}$).

2.2 Purification of LDH_{TT} and PK_{GST}

One Shot DE3 BL21 star *E. coli* cells (Invitrogen) were transformed with either the $pLDH_{TTHERMOPHILUS}$ or $pPK_{GSTEAROTHERMOPHILUS}$ vector. Cells were grown in LB media containing 50 μg/ml kanamycin to an approximate OD of 0.6 at 37℃. Induction was initiated by the addition of IPTG to 0.5 mM. After 6 hours, cells were pelleted by centrifugation at 10,000g for 30 minutes. The pellet was resuspended in buffer containing 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, and 30 mM Imidazole pH 8.0. Cells were lysed by sonication and cell debris removed subsequently by centrifugation at 15,000g.

HisBind resin (Novagen) was used to purify LDH_{TT} and PK_{GST} following the manufacturer's protocols (column volume = 10 ml). The supernatant from the lysed cells was loaded onto the resin and washed with 3 column volumes of wash buffer (20 mM Tris-HCl buffer pH 7.5, 1 M NaCl, 40 mM Imidazole pH 8.0). The protein was eluted with 4 column volumes of a buffer containing 20 mM Tris-HCl buffer pH 7.5, 0.5 M NaCl, 0.5 M imidazole pH 8.0, 0.2 mM PMSF, and 5 mM β-mercaptoethanol. Purity was estimated to be greater than 85% by SDS-PAGE. Fractions containing PK_{GST} of LDH_{TT} were pooled and dialyzed into 50 mM Tris-HCl buffer pH 7.5, 30 mM imidazole pH 8.0, and 150 mM NaCl. LDH_{TT} was dialyzed into a high salt storage buffer to reduce precipitation (50 mM Tris-HCl buffer pH 7.5, 150 mM Imidazole pH 8.0, and 500 mM NaCl). $AK_{(BSUB)}$ purification was purified as described previously [6]. Protein concentrations were estimated from the calculated extinction coefficients of 13,075 $M^{-1}cm^{-1}$ for PKGST and 27,515 $M^{-1}cm^{-1}$ for LDH_{TT}.

2.3 Characterization of PK_{GST} and LDH_{TT}

Temperature and substrate dependent activities of PK_{GST} and LDH_{TT} were determined over a range of temperatures. All assays were performed in Reaction Buffer (50 mM potassium phosphate buffer pH 7.5, 65 mM KCl and 5 mM $MgCl_2$). The Reaction Buffer for each concentration was incubated for 5 minutes at each temperature. The reaction rates of LDH_{TT} were measured at 45, 50, 60, 70, and 75℃ with 0.3 mM NADH for pyruvate concentrations of 1000, 750, 500, 100, 50, 10, and 5 μM respectively. To initiate the reaction LDH_{TT} was added after equilibration, and product formation (NAD^+) was quantitated spectrophotometrically at 340 nm. The initial reaction rate (v_0) was measured from the slope of the measured reaction curve for the first 6 seconds after placing the cuvette into the spectrophotometer. The reaction rates of PK_{GST} were measured at 45, 50, 60, 70, and 75℃ with 2.0 mM phosphoenolpyruvate (PEP), 0.3 mM NADH and ADP concentrations of 1000, 750, 500, 100, 50, 10, and 5 μM respectively. 0.5 μM LDHTT was added to the PK_{GST} assay mixture at 50, 60, 70, and 75℃. However, at temperatures below 50℃, 1.0 μM LDH_{TT} was required. PK_{GST} was added after the 5 minute incubation to initiate the reaction, and the rate of NADH consumed was measured at 340 nm. The plot of v_0 as a measure of pyruvate or ADP concentration for LDH_{TT} and PK_{GST} were fitted to the Michaelis-Menten equation to calculate the maximum velocity (V_{max}) and K_M using the software program Graphpad Prism.

2.4 Adenylate Kinase Protein Purification

Adenylate kinase was expressed in One Shot DE3 BL21 star *E. coli* transfromed with $pET11a_{AK_{(BSUB)}}$. 2 L of cells in LB media containing ampicillin at a concentration of

100 μg/ml were grown at 37℃ until and OD of 0.6 to 0.8 was reached. The media was then centrifuged at 10,000 rpm for 20 minutes to pellet the cells, and frozen overnight. The cells were resuspended in 40 ml resuspension buffer (buffer A). Buffer A contains 50 mM Tris buffer pH 7.5, 1 mM NaCl, 1mM MgCl, 0.1 mM EDTA, and 0.3 mM DDT. The cell pellet was thawed for 20 minutes on ice before sonication. Resuspended cells were sonicated for 4 times on output control 7, duty cycle 70, for 2 minutes on/2 minutes off schedule for a total of 8 minutes of sonication. The cells were then centrifuged at 16,000 rpm for 25 minutes at 6℃.

The first purification step was on a Q Sepharose HP column. Pressure limit was set to 0.5 MPa. The column was washed first with 1 column volume (CV) of buffer A, 50 ml total. A second was with 1 CV of a high salt buffer (buffer B) was run through the column. Buffer B is identical to buffer A except containing 1 M NaCl. Finally the column was equilibrated with 5 CV's of buffer A. The protein was filtered through a 0.22 μm filter before being loaded into the column. Sample was injected to 40 ml. A gradient from buffer A to 30% buffer B was run on the column, and the column flow was then held ad 30% buffer B. Column flow was fractionated into 5 ml volumes. The UV spectrum was used to identify fractions containing the protein.

Column fractions were pooled and the solution dialyzed into buffer A with 20 mM tris. The second step of purification was to run the protein through a blue sepharose affinity column. The column was washed with 10 CV's of ddH_2O (CV of 50 ml). A second wash step was done with 2 CV's buffer A. Dialyzed protein was concentrated to 5 to 10 ml and was added to the column. The column was washed with 8 CV's of buffer A until the OD from the UV spec at 280 nm was less than 0.1. Protein was eluted by using a mix from 0% to 100% buffer B (in buffer A) solution over 20 CV's. Efflux was fractionated into 5 ml fractions and the OD was measured and fractions

were collected until the OD was below 0.5. Fractions were run on a 17% SDS-page gel for 1:30 minutes at 80V and the fractions containing protein were pooled and dialyzed overnight into 20 mM pH 7.0 HEPES buffer.

A Superdex gel filtration column was used for the last purification step. The column was run with a buffer of 20 mM HEPES pH 7.4 and 0.2 M NaCl. A flow rate of 0.5 ml/min was run on the column. Volume fractions of 2 ml were collected and the column was run until the protein was fully eluted. The fractions containing protein were run on an SDS page gel and fractions containing protein were pooled. The protein was dialyzed into 20 mM HEPES overnight. AK was concentrated to an OD/ml of 30. Protein purification usually resulted in around 100 mg total protein. The extinction coefficient for $AK_{(BSUB)}$ used was 12,345 $M^{-1}cm^{-1}$.

2.5 Quantitation of Kinase Activity at 45-75°Cby the LDH_{TT}/PK_{GST} Coupled assay

All assays were performed in Reaction Buffer (50 mM potassium phosphate buffer pH 7.5, 65 mM KCl and 5 mM $MgCl_2$). The reaction rates of $AK_{(BSUB)}$ were measured at 1.4 mM AMP, 2 mM PEP, and 0.3 mM NADH for ATP concentrations of 1000, 750, 500, 100, 50, 10, and 5 μM respectively. The concentration of $AK_{(BSUB)}$ was 5 nM for all reactions. For the $AK_{(BSUB)}$ measurements the reaction buffer was incubated for 5 minutes to the correct temperature with 0.5 μM LDH_{TT} and 24 μM PK_{GST}. To reduce ATP hydrolysis ATP was added just prior to $AK_{(BSUB)}$to initiate the reaction. The initial reaction rate v_0 was measured in the same manner as for LDH_{TT} and PK_{GST}. The plot of v_0 vs ATP was fitted to the Michaelis-Menten equation in Graphpad PRISM to calculate V_{max} and K_M (Figure 3).

2.6 Numerical Simulation of the Coupled Assay

A numerical simulation using MATLAB was used to predict how the concentrations of the substrates varied during the reaction using single substrate and simple bi-substrate Michaelis-Menten kinetics. The reaction rates for PK, AK, and LDH were modeled using the Michaelis-Menten equation for a bi-substrate reaction with random substrate binding as described in equations 2.1, 2.2 and 2.3 respectively.

$$v_{AK} = \frac{v_{\max}[\mathrm{ATP}][\mathrm{AMP}]}{\mathrm{k_M^{ATP}k_M^{AMP}} + \mathrm{k_M^{AMP}}[\mathrm{ATP}] + \mathrm{k_M^{ATP}}[\mathrm{AMP}]} \tag{2.1}$$

$$v_{PK} = \frac{v_{\max}[\mathrm{ADP}][\mathrm{PEP}]}{\mathrm{k_M^{ADP}k_M^{PEP}} + \mathrm{k_M^{PEP}}[\mathrm{ADP}] + \mathrm{k_M^{ADP}}[\mathrm{PEP}]} \tag{2.2}$$

$$v_{LDH} = \frac{v_{\max}[\mathrm{Pyruvate}][\mathrm{NADH}]}{\mathrm{k_M^{Pyruvate}k_M^{NADH}} + \mathrm{k_M^{NADH}}[\mathrm{Pyruvate}] + \mathrm{k_M^{Pyruvate}}[\mathrm{NADH}]} \tag{2.3}$$

The rate of change in concentration of each of the products and substrates at time t can be calculated by using the enzyme specific activity v. ATP is the substrate of the AK reaction (2.1 and the product of the PK reaction (2.2).

$$\frac{\partial[\mathrm{ATP}]}{\partial t} = v_{PK} - v_{AK} \tag{2.4}$$

Similar equations exist for the other substrates AMP, ADP, PEP, Pyruvate, and NADPH.

$$\frac{\partial[\mathrm{AMP}]}{\partial t} = -v_{AK} \tag{2.5}$$

$$\frac{\partial[\mathrm{ADP}]}{\partial t} = v_{AK} - v_{PK} \tag{2.6}$$

$$\frac{\partial[\mathrm{PEP}]}{\partial t} = -v_{PK} \tag{2.7}$$

$$\frac{\partial[\text{Pyruvate}]}{\partial t} = v_{PK} - v_{LDH} \tag{2.8}$$

$$\frac{\partial[\text{ADP}]}{\partial t} = -v_{LDH} \tag{2.9}$$

A Euler approximation was used to calculate the solution of the series of partial differential equations with a step size of 10^{-3}. The experimental concentrations of the substrates and enzymes were used as the initial conditions for the numerical simulation. The initial values used were: AMP 1400 μM, ADP 0, PEP 2000 μM, Pyruvate 0, and NADH 300 μM. The values used for the kinetic parameters of AK were: V_{max} 16.03 μM/nM/s $\times$ 5 nM, ATP K_M 0.92 μM, AMP K_M 0.92 μM. For PK the values used were: V_{max} 0.68 μM/nM/s $\times$ 24000 nM, K_M 3026 μM. For LDH the values used were: V_{max} 1.71 μM/nM/s $\times$ 500 nM, Pyruvate K_M 142.7 μM, NADH K_M142.7 μM. The same K_M was used for both PEP and ADP for the PK reaction. A numerical solution for concentrations in the experiment was calculated for a total time period of 2 minutes after the reaction started. Numerical solutions for ATP concentrations corresponding to those used in the assay were calculated; 1000, 750, 500, 100, 50, 10, and 5 μM ATP.

2.7 Adenylate Kinase Evolution Experiment

Adaptation of *G. stearothermophilus* NUB3621-R:ThEV to higher temperature was done in a turbidostat system with a temperature ramp of 0.2 ℃. Bacteria were grown in LB supplemented with 0.59M $MgSO_4$, 0.91M CaCl, 0.04M $FeSO_4$, and 1.05M NTA. 5 mg/ml rifampicin and 10 mg/ml chloramphenicol were also added to the media. The media was prepared in 10 L batches, and cooled overnight before addition of the supplements and the antibiotics. The vessel volume of the turbidostat was

filled with modified LB media up to 400 ml. The turbidostat was filled and the stirring turned on, but the temperature was system was allowed to rest overnight to test for contamination before the temperature was increased. Glycerol stock of *G. stearothermophilus* NUB3621-R:ThEV AK_{Q199R} was plated onto modified LB plates, with rifampicin and chloramphenicol, and grown overnight at 60℃. A single colony was used to inoculate a 10 ml test tube with modified LB, plus rifampacin and chloramphenicol, and grown overnight at 60℃. The vessel was inoculated with the 10 ml test tube after allowing to equilibrate to 60℃. OD was measured continuously, and the control loop was turned on once the OD reached 0.2. Once the OD reached 0.2 the temperature was increased by 0.2℃, and the first sample was collected. Samples were collected every 12 hours, and the OD was measured at the same time to verify the turbidostat set-points. 50 ml 10% glycerol samples were frozen and stored in the -80℃ to restart the fermenter run in case the fermenter run was interrupted. A genomic DNA sample was prepared from a second 50 ml sample. A 1 ml glycerol sample was stored in the -80℃ for use in sequencing. The temperature was increased by 0.2℃ every 24 hours for 14 days to a final temperature of 62.4 ℃.

2.8 Derivation of Mathematical Equations

2.8.1 Derivation of Diffusion Equations

Diffusion across the membrane is governed by Fick's law (2.10)

$$J = -D\Delta\phi \tag{2.10}$$

The gradient $\Delta\phi$ is the difference in concentration across the membrane for both the outer membrane $[MCN_O]-\frac{1}{2}[MCN_P]$ and the inner membrane $[MCN_P]-\frac{1}{2}[MCN_C]$. Due to changes in pH the apparent concentration of MCN is 1/2 of the actual concentration since only certain MCN species can diffuse through each membrane. The diffusivity $-D$ is the permissivity of the membrane multiplied by the area of the membrane, $P_O A_O$. The diffusion across either membrane is therefore;

$$J_O = D_O \left([\mathrm{MCN}_O] - \frac{1}{2}[\mathrm{MCN}_P] \right) \tag{2.11}$$

$$J_I = D_I \left([\mathrm{MCN}_P] - \frac{1}{2}[\mathrm{MCN}_C] \right) \tag{2.12}$$

The flux across each membrane into the periplasm per unit volume is;

$$J_O = \frac{P_O A_O}{V_P} \left([\mathrm{MCN}_O] - \frac{1}{2}[\mathrm{MCN}_P] \right) \tag{2.13}$$

$$J_I = \frac{P_I A_I}{V_C} \left([\mathrm{MCN}_P] - \frac{1}{2}[\mathrm{MCN}_C] \right) \tag{2.14}$$

At steady state the concentration inside the periplasm is constant so the flux across the outer and inner membranes is equal; $\frac{\partial[\mathrm{MCN}_O]}{\partial t} = \frac{\partial[\mathrm{MCN}_I]}{\partial t}$.

$$\frac{P_O A_O}{V_P} \left([\mathrm{MCN}_O] - \frac{1}{2}[\mathrm{MCN}_P] \right) = \frac{P_I A_I}{V_C} \left([\mathrm{MCN}_P] - \frac{1}{2}[\mathrm{MCN}_C] \right) \tag{2.15}$$

replace $\frac{P_I A_I}{V_I}$ with D_I.

$$\frac{D_O}{D_I} \left([\mathrm{MCN}_O] - \frac{1}{2}[\mathrm{MCN}_P] \right) = \left([\mathrm{MCN}_P] - \frac{1}{2}[\mathrm{MCN}_C] \right)$$

$$\frac{D_O}{D_I}[\mathrm{MCN}_O] - \frac{P_O A_O}{P_I A_I}\frac{1}{2}[\mathrm{MCN}_P] = [\mathrm{MCN}_P] - \frac{1}{2}[\mathrm{MCN}_C]$$

$$\frac{D_O}{D_I}[\mathrm{MCN}_O] + \frac{1}{2}[\mathrm{MCN}_C] = \frac{D_O}{D_I}\frac{1}{2}[\mathrm{MCN}_P] + [\mathrm{MCN}_P]$$

$$[\mathrm{MCN}_P] = \frac{[\mathrm{MCN}_C]/2 + \frac{D_O}{D_I}[\mathrm{MCN}_O]}{1 + \frac{D_O}{D_I}\frac{1}{2}}$$

The flux through the inner membrane is therefore

$$J_I = \frac{P_I A_I}{V_I}\left(\frac{[\mathrm{MCN}_C]/2 + \frac{D_O}{D_I}[\mathrm{MCN}_O]}{1 + \frac{D_O}{D_I}\frac{1}{2}} - \frac{1}{2}[\mathrm{MCN}_C]\right) \tag{2.16}$$

2.8.2 Derivation of Kinetic Equations

TetX2 catalyzes the degradation of minocycline through a random binding bisubstrate reaction which requires NADPH and minocycline. The reaction rate V_{max} can be calculated using Michealis-Menten bisubstrate kinetics.

$$v = \frac{\frac{V_{max}[\mathrm{NADPH}]}{K_M^{\mathrm{NADPH}} + [\mathrm{NADPH}]}[\mathrm{MCN}_C]}{\frac{K_M^{\mathrm{MCN}}[\mathrm{NADH}]}{K_M^{\mathrm{NADPH}} + [\mathrm{NADPH}]} + [\mathrm{MCN}_C]} \tag{2.17}$$

V_{max}obs and $K_M$obs can be used to simplify the bisubstrate equation.

$$V_{max}^{\mathrm{obs}} = \frac{V_{max}[\mathrm{NADPH}]}{K_M^{\mathrm{NADPH}} + [\mathrm{NADPH}]} \tag{2.18}$$

$$K_M^{\mathrm{obs}} = \frac{K_M^{\mathrm{MCN}}[\mathrm{NADPH}]}{K_M^{\mathrm{NADPH}} + [\mathrm{NADPH}]} \tag{2.19}$$

$$v = \frac{V_{max}^{\mathrm{obs}}[\mathrm{MCN}]}{K_M^{\mathrm{obs}} + [\mathrm{MCN}]} \tag{2.20}$$

The degradation rate of minocycline in the cell is calculated by the reaction speed of TetX2 multiplied by the concentration of TetX2 and divided by the volume of the cell.

$$V = \frac{V_{max}^{\mathrm{obs}}[\mathrm{TetX2}][\mathrm{MCN}]}{K_M^{\mathrm{obs}} + [\mathrm{MCN}]} \tag{2.21}$$

2.8.3 Derivation of Minocycline Concentration

At steady state the influx of minocycline (equation 2.16) through the inner membrane is equal to the degradation rate of TetX2 (equation 2.12);

$$\frac{V_{max}^{\text{obs}}[\text{TetX2}][\text{MCN}_C]}{K_M^{\text{obs}} + [\text{MCN}_C]} = D_I \left([\text{MCN}_P] - \frac{1}{2}[\text{MCN}_C] \right) \tag{2.22}$$

$$\frac{V_{max}^{\text{obs}}[\text{TetX2}][\text{MCN}_C]}{K_M^{\text{obs}} + [\text{MCN}_C]} = D_I \left([\text{MCN}_P] - \frac{1}{2}[\text{MCN}_C] \right)$$

$$\frac{\frac{V_{max}^{\text{obs}}}{D_I}[\text{TetX2}][\text{MCN}_C]}{K_M^{\text{obs}} + [\text{MCN}_C]} = [\text{MCN}_P] - \frac{1}{2}[\text{MCN}_C]$$

$$\frac{V_{max}^{\text{obs}}}{D_I}[\text{TetX2}][\text{MCN}_C] = \left([\text{MCN}_P] - \frac{1}{2}[\text{MCN}_C] \right) \left(K_M^{\text{obs}} + [\text{MCN}_C] \right)$$

$$\frac{V_{max}^{\text{obs}}}{D_I}[\text{TetX2}][\text{MCN}_C] = [\text{MCN}_P] \left(K_M^{\text{obs}} + [\text{MCN}_C] \right) - 0.5[\text{MCN}_C]K_M^{\text{obs}} - 0.5[\text{MCN}_C]^2$$

$$[\text{MCN}_P] \left(K_M^{\text{obs}} + [\text{MCN}_C] \right) = \frac{V_{max}^{\text{obs}}}{D_I}[\text{TetX2}][\text{MCN}_C] + 0.5[\text{MCN}_C]K_M^{\text{obs}} + 0.5[\text{MCN}_C]^2$$

$$[\text{MCN}_P] \left(K_M^{\text{obs}} + [\text{MCN}_C] \right) = \left(\frac{V_{max}^{\text{obs}}}{D_I}[\text{TetX2}] + 0.5K_M^{\text{obs}} \right) [\text{MCN}_C] + 0.5[\text{MCN}_C]^2$$

$$[\text{MCN}_P] = \frac{\left(\frac{V_{max}^{\text{obs}}}{D_I}[\text{TetX2}] + 0.5K_M^{\text{obs}} \right) [\text{MCN}_C] + 0.5[\text{MCN}_C]^2}{K_M^{\text{obs}} + [\text{MCN}_C]}$$

replace $[\text{MCN}_P]$ with equation (2.19).

$$\frac{[\text{MCN}_C]/2 + \frac{D_O}{D_I}[\text{MCN}_O]}{1 + 0.5\frac{D_O}{D_I}} = \frac{\left(\frac{V_{max}^{\text{obs}}}{D_I}[\text{TetX2}] + \frac{1}{2}K_M^{\text{obs}} \right) [\text{MCN}_C] + 0.5[\text{MCN}_C]^2}{K_M^{\text{obs}} + [\text{MCN}_C]}$$

replace with $1 + 0.5\frac{D_O}{D_I}$ with D_P.

$$\frac{[\mathrm{MCN}_C]/2 + \frac{D_O}{D_I}[\mathrm{MCN}_O]}{D_P} = \frac{\left(\frac{V_{max}^{\mathrm{obs}}}{D_I}[\mathrm{TetX2}] + \frac{1}{2}K_M^{\mathrm{obs}}\right)[\mathrm{MCN}_C] + 0.5[\mathrm{MCN}_C]^2}{K_M^{\mathrm{obs}} + [\mathrm{MCN}_C]}$$

$$\left([\mathrm{MCN}_C]/2 + \frac{D_O}{D_I}[\mathrm{MCN}_O]\right)\left(K_M^{\mathrm{obs}} + [\mathrm{MCN}_C]\right) = \left(\left(\frac{V_{max}^{\mathrm{obs}}}{D_I}[\mathrm{TetX2}] + \frac{1}{2}K_M^{\mathrm{obs}}\right)[\mathrm{MCN}_C] + 0.5[\mathrm{MCN}_C]^2\right) D_P$$

$$0.5[\mathrm{MCN}_C]K_M^{\mathrm{obs}} + \frac{D_O}{D_I}[\mathrm{MCN}_O]K_M^{\mathrm{obs}} + \frac{D_O}{D_I}[\mathrm{MCN}_O][\mathrm{MCN}_C] + 0.5[\mathrm{MCN}_C]^2 = \left(\frac{V_{max}^{\mathrm{obs}}}{D_I}[\mathrm{TetX2}] + \frac{1}{2}K_M^{\mathrm{obs}}\right)[\mathrm{MCN}_C]D_P + 0.5[\mathrm{MCN}_C]^2 D_P$$

$$0 = 0.5[\mathrm{MCN}_C]K_M^{\mathrm{obs}} + \frac{D_O}{D_I}[\mathrm{MCN}_O]K_M^{\mathrm{obs}} + \frac{D_O}{D_I}[\mathrm{MCN}_O][\mathrm{MCN}_C] + 0.5[\mathrm{MCN}_C]^2 - \left(\frac{V_{max}^{\mathrm{obs}}}{D_I}[\mathrm{TetX2}] + \frac{1}{2}K_M^{\mathrm{obs}}\right)[\mathrm{MCN}_C]D_P - 0.5[\mathrm{MCN}_C]^2 D_P$$

$$\begin{aligned} 0 = {} & \frac{D_O}{D_I}[\mathrm{MCN}_O]K_M^{\mathrm{obs}} \\ & + \left(\frac{D_O}{D_I}[\mathrm{MCN}_O] + 0.5K_M^{\mathrm{obs}} - 0.5D_P K_M^{\mathrm{obs}} - \frac{D_P}{D_I}V_{max}^{\mathrm{obs}}[TetX2]\right)[\mathrm{MCN}_C] \\ & + (1 - D_P)\,0.5[\mathrm{MCN}_C]^2 \end{aligned}$$

$$0 = D_O[\mathrm{MCN}_O]K_M^{\mathrm{obs}}$$
$$+ \left(D_O[\mathrm{MCN}_O] + (D_I - D_I D_P)0.5K_M^{\mathrm{obs}} - D_P V_{max}^{\mathrm{obs}}[TetX2]\right)[\mathrm{MCN}_C]$$
$$+ (D_I - D_I D_P)\,0.5[\mathrm{MCN}_C]^2$$

Replacing D_P with $1 + 0.5\frac{D_0}{D_I}$;

$$D_I - D_I D_P = D_I - D_I\left(1 + 0.5\frac{D_O}{D_I}\right)$$

$$D_I - D_I D_P = \cancel{D_I} - \cancel{D_I} - 0.5\frac{\cancel{D_I} D_O}{\cancel{D_I}}$$

$$D_I - D_I D_P = 0.5 D_O$$

$$0 = D_O[\mathrm{MCN}_O]K_M^{\mathrm{obs}}$$
$$+ \left(D_O[\mathrm{MCN}_O] + D_O 0.25K_M^{\mathrm{obs}} - D_P V_{max}^{\mathrm{obs}}[TetX2]\right)[\mathrm{MCN}_C]$$
$$+ D_O 0.5[\mathrm{MCN}_C]^2$$

Divide by D_O.

$$0 = [\mathrm{MCN}_O]K_M^{\mathrm{obs}}$$
$$+ \left([\mathrm{MCN}_O] + 0.25K_M^{\mathrm{obs}} - \frac{D_P}{D_O}V_{max}^{\mathrm{obs}}[TetX2]\right)[\mathrm{MCN}_C]$$
$$+ 0.25[\mathrm{MCN}_C]^2$$

$$0 = [\mathrm{MCN}_O]K_M^{\mathrm{obs}} + \left([\mathrm{MCN}_O] + 0.25K_M^{\mathrm{obs}} - \frac{1 - 0.5\frac{D_O}{D_I}}{D_O}V_{max}^{\mathrm{obs}}[TetX2]\right)[\mathrm{MCN}_C] + 0.25[\mathrm{MCN}_C]^2$$

$$\frac{1 - 0.5\frac{D_O}{D_I}}{D_O} = \frac{D_I - 0.5D_O}{D_O D_I}$$

$$\frac{1 - 0.5\frac{D_O}{D_I}}{D_O} = \frac{1}{D_O} - 0.5\frac{1}{D_I}$$

$$0 = [\mathrm{MCN}_O]K_M^{\mathrm{obs}} + \left([\mathrm{MCN}_O] + 0.25K_M^{\mathrm{obs}} - \left(\frac{1}{D_O} - 0.5\frac{1}{D_I}\right)V_{max}^{\mathrm{obs}}[TetX2]\right)[\mathrm{MCN}_C] + 0.25[\mathrm{MCN}_C]^2$$

We can replace $\frac{1}{D_O} - 0.5\frac{1}{D_I}$ with $\frac{1}{D}$ since the diffusion constant will be a variable.

$$0 = [\mathrm{MCN}_O]K_M^{\mathrm{obs}} + \left([\mathrm{MCN}_O] + 0.25K_M^{\mathrm{obs}} - \frac{V_{max}^{\mathrm{obs}}}{D}[TetX2]\right)[\mathrm{MCN}_C] + 0.25[\mathrm{MCN}_C]^2 \quad (2.23)$$

Solving the quadratic equation for MCN yields;

$$a = -0.25$$

$$b = [\mathrm{MCN}_O] - 0.25K_M^{\mathrm{obs}} - \frac{V_{max}^{\mathrm{obs}}}{D}[TetX2]$$

$$c = K_M^{\mathrm{obs}}[\mathrm{MCN}_O]$$

$$[\mathrm{MCN}_C] = 2\left([\mathrm{MCN}_O] - 0.25K_M^{\mathrm{obs}} - \frac{V_{max}^{\mathrm{obs}}}{D}[TetX2]\right) - 2\sqrt{\left([\mathrm{MCN}_O] - 0.25K_M^{\mathrm{obs}} - \frac{V_{max}^{\mathrm{obs}}}{D}[TetX2]\right)^2 + K_M^{\mathrm{obs}}[\mathrm{MCN}_O]} \quad (2.24)$$

$$GR = 1 - \frac{[\mathrm{MCN}_C]^B}{A + [\mathrm{MCN}_C]^B} \quad (2.25)$$

2.9 Measuring Population Dynamics Using Next Generation Sequencing

We used a newly developed barcoding method called FREQ-SEQ [Chubiz et al., 2012] to generate population distribution histograms for TetX2 evolution experiments. Genomic DNA from each the heterogenous populations was isolated from each day of the experiment. An initial sample was taken as an initial timepoint, for a total of 31 samples covering the 10 experiments for 3 days. Genomic DNA was isolated using a Mo Bio UltraClean Microbial DNA Isolation kit. An initial PCR step was used to amplify the region the *tetx* gene is in.

Primer pair for genome amplification:

Forward: 5GAAGAAGGCGAACCTGAAAGGCTACGGCCGATAATTGGTCGCC-CGAGAAGTAGAGGACCTAGAAGGAGAC

Reverse: 5CGGCATAATTTCTTGACGGAAGCACGAACTTTCATTTTTACTCTC-CGTAACTTATTATACATTTAACAATTGCTG

The barcoded primers were amplified off of a plasmid derivative of pUC19 stored in DH10B *E. coli*cells in glycerol stocks at -80℃ (gift of C. Marx, Harvard).

Illumina barcode amplification primers:

Forward: AATGATACGGCGACCAC

Reverse: ACTGGCCGTCGTTTTAC

Gel extraction and ethanol precipitation were used to purify the barcoded primers. A 10% acrylamide gel was run for 30 minutes to purify the PCR solution. The acrylamide gel was stained with a 10% ethidium bromide solution for 30 minutes. The band containing the barcode primer was then cut out of the gel. Gel electrophoresis was used to extract the DNA from the gel fragment. The gel fragment was placed inside a dialysis bag, with 12,000 to 14,000 MWCO, with 300 ml of buffer and the bag was placed inside the gel box antiparallel to the running direction. Electrophoresis was run for 30 minutes to allow the barcode DNA fragment to run out of the gel fragment into the dialysis bag. 700 ml of ethanol were added to the solution after the contents of the dialysis bag were transferred to a 1.5 ml microcentrifuge tube. 10% ammonium acetate was then added to the solution. The solution was incubated overnight at -40℃. A second incubation was performed for an additional 1 hour in a dry ice ethanol bath. The bath needs to be kept at a temperature near -50℃ to prevent the solution from freezing. Centrifugation at 15,000 rpm in a microcentrifuge was performed for 30 minutes at 4℃ to pellet the DNA. The solution was then decanted and the precipitate lyophilized to remove all traces of ethanol contamination. The DNA was resuspended in 20 μl of ddH_2O.

Because the read length of Illumina sequencing is only around 100 bases the targeted base pair has to be near the start of the oligonucleotide used for barcoding. The gene regions surrounding each of the targeted base pairs were amplified from the product of the initial PCR reaction. The primer is also used to insert a segment of M13 DNA that is complementary to the barcode primer M13 region.

Illumina Forward primers: K64/A161:

Forward: GTAAAACGACGGCCAGTTGGTGGAACCCTTGACCTACA

Reverse: CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCCAC
TTCTTCTTACCAGGTTCAAGC

F235/T674:

Forward: GTAAAACGACGGCCAGTAACCCCAATAATAATGGTGCATTGCA
Reverse: CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCCCAATCA
TTGTTATGGGTAATGGG

T280/A808:

Forward: GTAAAACGACGGCCAGGACGAACGCTACAAAGAATTGATTC
Reverse: CAAGCAGAAGACGGCATACGAGCTCTTCCGATCCAATGCTA
TTAAATTTTCCATCGGC

S326/G947:

Forward: GTAAAACGACGGCCAGTCTTTTGCAGGGCAGGGAGTAAA
Reverse: CAAGCAGAAGACGGCATACGAGCTCTTCCGATCCAATTG
CTGAAACGTAAAGTCGGG

N371/A1082:

Forward: GTAAAACGACGGCCAGCAAAGAAGCACAAGAAGAATCAACTCAA
Reverse: CAAGCAGAAGACGGCATACGAGCTCTTCCGATCGCACAGACATA
CAAAAGATTGGC

Once the template strand and the barcode strand have been isolated the final step is to amplify both strands together, A third set of primers is used for the final extension reaction. The forward primer is complementary to the barcode strand, and the reverse primer is complementary to the template strand. The template and barcode strands contain the complementary M13 sequences to complete the extension. The final template is purified using a PCR purification kit. The 31 samples were mixed and then sent out for Illumina sequencing.

Chapter 3

Identification and Characterization of Novel Adenylate Kinase Mutants

3.1 Introduction

Chromosomal replacement of an essential gene with a maladapted copy focuses adaptation on our target gene by creating an enzymatic "weak-link" with reduced function. The "weak link" adaptation experiment is a powerful model system for exploring evolution because of the strong and precise correlation of activity to fitness. Success in the environment is strongly affected by function because the enzyme is essential for cell survival [Atkinson and Walton, 1967, Couñago and Shamoo, 2005]. Selection for mutations using the "weak link" method is precise because the system favors mutations that restore activity and provide the greatest fitness benefits. We used a previously created strain of *G. stearothermophilus* where the gene for adenylate kinase

has been replaced with the homologue from the mesophile *Bacillus subtilis* [Couñago and Shamoo, 2005]. Comparisons of the activity and stability of mutations that successfully restore AK mutants with ancestors provide insights into how fitness is influenced by enzymatic activity. We changed the selection strength of the adaptation experiment to find novel AK mutants that would allow us to compare the effects of selection strength on the accessible adaptive mutants.

Reducing selection strength alleviates constraints on possible adaptive pathways. Mutations that provide a moderate increase to fitness are allowed to persist in the population when selection is relaxed. New adaptive pathways and fitness peaks are revealed by increasing the number of allowed mutations at each step. We reduced the rate of temperature increase in the turbidostat to change the selection strength during adaptation of our "weak link" strain of *G. stearothermophilus*.

We started the adaptation experiment with the first mutant identified, $AK_{(BSUB)Q199R}$, to reduce the time required for the experiment. Since $AK_{(BSUB)Q199R}$ is fixed in the population until the temperature reaches 60℃ we started at that temperature to better capture the point in adaptation of *adk* where we see the most diversity of mutants. Using a temperature ramp of 0.2℃ per day compared with the previously used ramp of 0.5℃ altered the results of the turbidostat evolution experiment. We were able to identify two novel mutants of $AK_{(BSUB)Q199R}$ during two different adaptation experiments. The mutants that we discovered contained combinations of mutations that we had previously identified. These mutants were $AK_{(BSUB)Q199R/T179I/Q16L}$ and $AK_{(BSUB)Q199R/A193V/Q16L}$. Once new mutations were identified our goal was to characterize the mutants to compare them with the mutants isolated from the previous experiment. We developed a thermostable coupled assay to better measure the activity of the novel mutants and compare them to previously identified mutations.

The pyruvate kinase (PK: 2.7.1.40) and lactate dehydrogenase (LDH: 1.1.1.27) coupled assay is a well described method for measuring activity of enzymes catalyzing phosphotransferase reactions involving the generation of ADP at mesophilic temperatures [Kornberg and Pricer Jr, 1951]. However, use of the coupled assay to measure AK activity is limited to temperatures lower than 45℃. Many thermophilic enzymes have very low activity at these reduced temperatures making meaningful measurements difficult. In addition, extrapolating kinetic measurements obtained at 20 to 40℃ to much higher temperatures leads to significant propagation errors that undermine the quality of interpretation. In the coupled assay, kinase activity is measured indirectly by coupling the formation of ADP to the oxidation of NADH to NAD^+ (Figure 3.1) [Saint Girons et al., 1987]. To circumvent the limited range of the assay, an end point coupled assay, using mesophilic pyruvate kinase (PK) and L-lactate dehydrogenase (LDH), has been used previously to characterize the activity of thermophilic kinases [Couñago et al., 2008] [Glaser et al., 1992]. The phosphotransferase reaction is carried out at the desired high temperature and is then quenched at specific time intervals. Afterwards the PK and LDH reactions are performed at 37℃. However, unlike the continuous assay, this two-step assay is very time consuming and relies on accurately quenching the reactions at each time point to build a reaction curve [Peña et al., 2011].

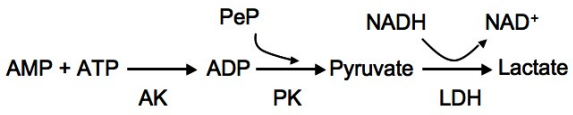


Figure 3.1: Coupled Assay for Adenylate Kinase

An often-overlooked attribute of the coupled assay, whether done in an end-point

or continuous fashion, is the role of ATP feedback. Since the ADP used by PK regenerates ATP, creating a positive feedback loop, appropriate conditions for the assay are important to produce meaningful data. This is particularly true when the assay is being used at high temperatures, where rates would be markedly higher and the linear range of the assay accordingly worse. Interestingly, the caveats associated with the use of this assay have not been considered previously, leading us to critically evaluate the feedback effects on the observed rates of NAD^+ production.

The reaction rate curve can be measured continuously, simplifying measurement of kinetic parameters for the target enzyme. Adenylate Kinases from *Bacillus subtilis* ($AK_{(BSUB)}$) that had been evolved to higher thermostability in a previously published report were used to test the efficacy of the assay [Couñago et al., 2008]. Results were compared to a previous study in the lab done by Matthew Peña, in which $AK_{(BSUB)}$ activity was measured with an end point assay [Peña et al., 2011]. The important role of feedback was addressed through mathematical modeling and experimental validation.

3.2 LDH_{TT} and PK_{GST} have linear responses to substrate production from 45-75℃

The activities of LDH_{TT} and PK_{GST} at 45, 50, 60, 70, and 75℃ were measured to determine the working range for the high temperature coupled assay (tables 3.2 and 3.3). Although the reaction rate naturally decreases with temperature, the reaction rates for the two enzymes remain fast enough that the limiting reaction is still the adenylate kinase reaction. LDH_{TT} activity decreases such that at less than 45℃, overall activity is too low to effectively measure AK activity. Conversely, PK_{GST}

Table 3.1: LDH_{TT} activity at NADH Concentrations of 300 μM and 150 μM

NADH (μM)	k_{cat} (μMs^{-1})	K_M(μM)	k_{cat}/K_M (s^{-1})
300	0.08 ± 0.00	142.7 ± 21.31	0.03 ± 0.01
150	0.09 ± 0.01	155.7 ± 57.10	0.03 ± 0.01

Table 3.2: LDH_{TT} kinetics at 45 to 70℃

T (℃)	k_{cat} (μMs^{-1})	K_M(μM)
75℃	0.1 ± $4.5e^{-3}$	98.3 ± 26.1
70℃	0.1 ± $4.2e^{-3}$	121.9 ± 13.7
60℃	0.08 ± $1.0e^{-3}$	142.7 ± 21.3
50℃	0.03 ± $1.0e^{-3}$	217.4 ± 26.0
45℃	0.02 ± $3.5e^{-3}$	195.8 ± 57.1

LDH_{TT} was measured at 72.7 nM for all reactions.

Table 3.3: PK_{GST} kinetics at 45 to 70℃

T (℃)	k_{cat} (μMs^{-1})	K_M(μM)
75℃	2.1 ± $6.5e^{-4}$	496 ± 360
70℃	1.1 ± $1.0e^{-4}$	563 ± 115
60℃	1.0 ± $4.7e^{-5}$	326 ± 43
50℃	6.0 ± $4.0e^{-5}$	325 ± 62
45℃	7.7 ± $8.3e^{-5}$	688 ± 151

PK_{GST} was measured at 1.6 μM for 75℃and 8 μM for all other temperatures.

starts to precipitate at 80℃, presumably due to denaturation. The assay is therefore best used in the range of 45 to 75℃. LDH_{TT} was also assayed at a concentration of 150 μM NADH to compare the assay's response to changes in initial NADH concentration. k_{cat}/K_M were in good agreement at 150 and 300 μM NADH (table 3.1). Although the AK reaction is limiting, there can be a lag phase in the reaction curve during which the substrates for the two subsequent reactions accumulate, undermining the quality of observed kinetic parameters. To minimize the lag phase, the reaction rates of LDH_{TT} and PK_{GST} should exceed the maximum reaction rate of AK at least 6 fold [Easterby, 1981].

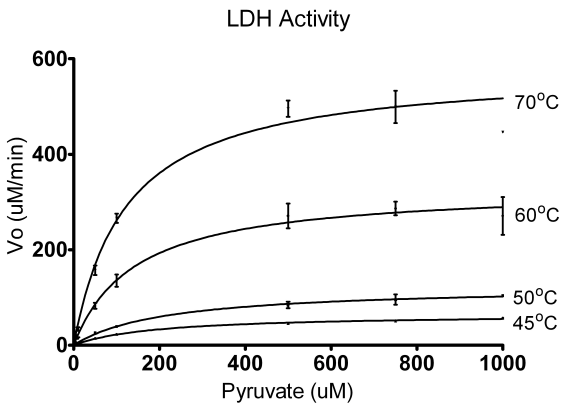


Figure 3.2: LDH_{TT} at 45 to 75℃ shows a linear increase in activity with respect to temperature above 45℃, below which LDH activity decreases significantly. Reactions were measured in triplicate at 340 nM for 2 minutes, and the first 6 seconds of reaction time were used to measure the initial velocity. The reaction rate curves were fit to Michaelis-Menten kinetics.

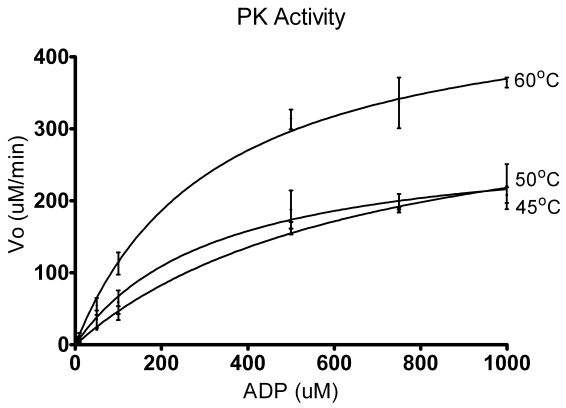


Figure 3.3: PK_{GST} shows a robust response range from 45 to 60℃. While PK is still active at 80℃, the protein aggregates at this temperature. Reactions were measured in triplicate at 340 nm for 2 minutes. The first 6 seconds of reaction time were used to measure the initial velocity. The reaction rate curves were fit to Michaelis-Menten kinetics.

3.3 Modeling ATP feedback with a numerical simulation

ATP is not depleted in the coupled assay because it is a product of the PK reaction. The reaction rate increases once ATP starts to accumulate because ATP is a substrate of the AK reaction. A mathematical model using Michaelis-Menten kinetics was used to model the substrate concentrations of ATP, AMP, ADP, pyruvate, and NADH to predict what time period effectively measures the reaction rate while minimizing the effect of ATP feedback. The reaction rates for AK cannot be modeled as single substrate reactions since AMP is depleted normally and begins to limit the reaction once ATP accumulates. The reaction rates for LDH_{TT} were also modeled using bi-substrate kinetics since pyruvate continues to be produced until ATP is depleted. PK_{GST} was modeled using a bi-substrate reaction because ADP concentration accumulates above PEP concentration before the reaction is complete. The concentration of ATP begins to increase significantly after more than 10 seconds at all initial ATP concentrations. The initial reaction rate was measured from the first 6 seconds of the reaction curve to minimize the effect of ATP feedback. Figure 3.4 shows the predicted changes in concentration for AMP, ATP, ADP, pyruvate, and NADH at 1000 μM ATP and 60℃ for 0.5 nM $AK_{(BSUB)Q199R/T179I}$. Experimentally measured values of V_{max} and K_M for LDH_{TT} and PK_{GST} (tables 3.2 and 3.3) were used for the simulation. Another concern when using this assay is the spontaneous conversion of ATP to ADP at higher temperatures. The increased rate of hydrolysis of ATP to ADP above 60℃ means ADP will start to accumulate, initiating the PK and LDH reactions before the target enzyme is even added to the reaction. Addition of ATP after temperature equilibration of the buffer limits hydrolysis from affecting

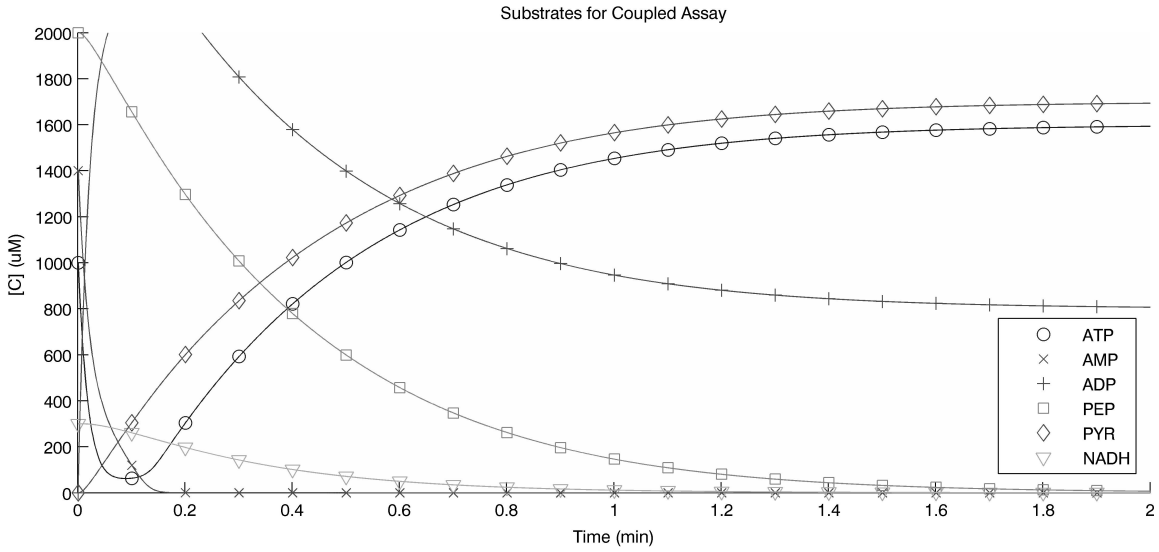


Figure 3.4: Mathematical simulation of concentrations of ATP, AMP, ADP, Pyruvate, and NADH at 1000 μM ATP show that ATP concentration starts to increase after ADP and pyruvate begin accumulating from the PK reaction. The reaction was simulated for 2 minutes using equations 2.1, 2.2 and 2.3 to determine the best range for measurement of kinase activity when feedback is present. We determined that the ATP concentration does not begin to increase until after more than 10 seconds of reaction time. Values for V_{max} and K_M are for $AK_{(BSUB)Q199R/T179I}$ from table 3.4 at 60℃ and 0.5 nM AK. Parameters for LDH_{TT} and PK_{GST} used values for V_{max} and K_M from tables 3.2 and 3.3 at 60℃. A concentration of 0.5 μM LDH_{TT} and 24 μM PK_{GST} was used to calculate the V_{max} from the measured k_{cat} in the simulation.

the reaction rate in the time frame of the experiment.

3.4 Comparison of data from the coupled and end point methods

The activity of a thermostable mutant of $AK_{(BSUB)Q199R/T179I}$ was measured at 50, 55, 60, and 65℃ (table 3.4). A second thermostable mutant $AK_{(BSUB)Q199R/G213E}$ was also measured at 60℃ (table 3.5). The calculated K_M and k_{cat} were compared to values from a previous study, done using an end point coupled assay where the

Table 3.4: $AK_{(BSUB)Q199R/T179I}$ kinetics at 55 to 70℃

T (℃)	k_{cat} (μMs^{-1})	K_M(μM)	K_M^{obs}(μM)	k_{cat}/K_M(s^{-1})
	Coupled Assay			
50℃	8.5 ± 0.7	17.0 ± 8.3	2.4 ± 1.2	3.5 ± 1.7
55℃	16.7 ± 0.6	10.9 ± 2.2	1.2 ± 0.2	13.8 ± 2.8
60℃	15.3 ± 0.2	24.2 ± 2.1	0.7 ± 0.1	21.1 ± 1.8
65℃	25.0 ± 1.6	253.1 ± 52.4	1.4 ± 0.3	18.0 ± 3.9
	End Point Assay [Peña et al., 2011]			
55℃	9.9 ± 0.2	16.2 ± 0.6	1.8 ± 0.2	5.5 ± 0.8
60℃	13.7 ± 0.2	23.4 ± 0.2	0.7 ± 0.1	19.6 ± 2.8
65℃	17.1 ± 1.2	163.8 ± 0.4	0.9 ± 0.2	19.0 ± 3.8

$AK_{(BSUB)}$ was measured at 5 nM for all temperatures.

AK reaction was decoupled from the PK and lactate dehydrogenase reactions [Peña et al., 2011]. The results from this previous study were used to ensure the accuracy of the high temperature coupled assay. At increased temperatures, there is potential for a significant loss of kinase activity due to unfolding, such that the observed K_M (K_M^{obs}) is often markedly different than the actual K_M [Peña et al., 2011]. Since the relevant temperature dependent unfolding properties of $AK_{(BSUB)Q199R/T179I}$ and $AK_{(BSUB)Q199R/G213E}$ are known, we are able to calculate K_M from the equation $K_M^{obs} = K_M(1 + K_{unf})$ (table 3.4).

3.5 Conclusions

PK_{GST} and LDH_{TT} isolated from thermophiles can be used to measure the rate of kinases such as $AK_{(BSUB)}$ by coupling the reduction of NADH to the production of

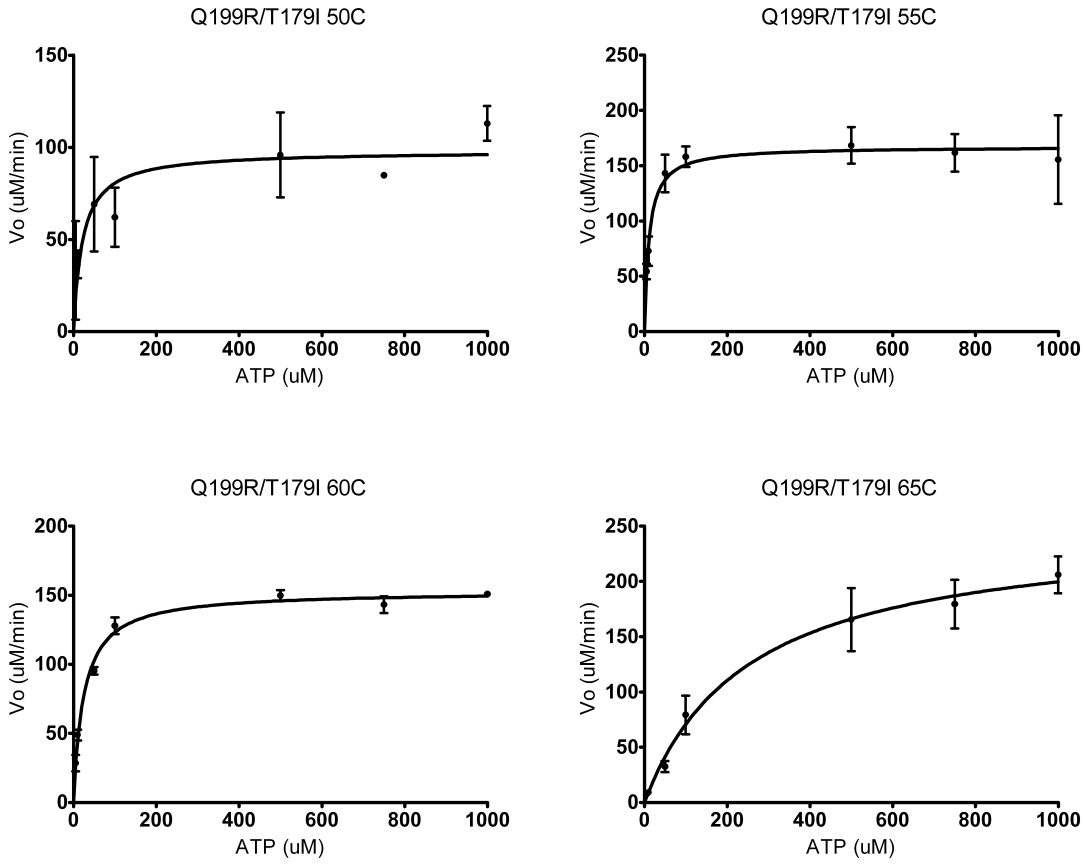


Figure 3.5: Activity of $AK_{(BSUB)Q199R/T179I}$ measured with the coupled assay is in good agreement with previous results using a more labor intensive end point assay [Peña et al., 2011]. $AK_{(BSUB)Q199R/T179I}$ concentration of 5 nM was used with 0.5 μM LDH_{TT} and 24 μM PK_{GST} except at 45℃ where 1.0 μM LDH_{TT} was used. AMP concentration was 1.4 mM. Reactions were measured in triplicate at 340 nM for 2 minutes, and the first 6 seconds of reaction time were used to measure the initial velocity. The reaction rate curves were fit using Michealis-Menten kinetics.

Table 3.5: Comparison of coupled assay and end point assay for $AK_{(BSUB)Q199R/G213E}$ at 60℃

	k_{cat} (μMs^{-1})	K_M(μM)	K_M^{obs}(μM	$k_{cat}/K_M(s^{-1})$
Coupled Assay	23.1 ± 0.8	0.5 ± 0.1	20.31 ± 4.0	42.6 ± 8.6
End Point Assay	11.2 ± 0.3	1.0 ± 0.2	37.5 ± 0.4	11.6 ± 2.4

ADP. This coupled assay is a fast and reliable method for measuring the biophysical parameters of AK and similar enzymes. However, drawbacks include the accumulation of ATP caused by feedback from the secondary PK reaction when the formation of ADP is being assayed. However, the assay can still be used when feedback is present as long as v_0 is measured from the initial reaction rate, before ATP begins to accumulate in significant amounts. In reactions where the ATP concentration is saturating, feedback is not a problem and v_0 can be measured from the linear phase of the reaction curve. We have determined the required concentrations of the secondary enzymes PK_{GST} and LDH_{TT} to match the expected rate of the measured enzyme at the target temperatures. The assay works best for enzyme concentrations that result in a V_{max} of less than 200 μM/s to allow for accurate measurement of the slope for v_0. At high temperatures, this assay is superior to kinase assays that utilize antibodies because the affinity of antibodies can vary with temperature; the coupled assay, however, is robust to small changes in secondary enzyme activity, as long as the reaction rate for the measured enzyme remains limiting. The accuracy of the assay is comparable to that of the end point assay previously used to measure thermophilic AK activity. PK_{GST} from *G. stearothermophilus* and LDH_{TT} from *T. thermophilus* can be used from 45 to 75℃ to measure the rate of phosphotransferase reactions that produce ADP. LDH and PK isolated from rabbit muscle can be used

to measure kinase activity at temperatures below 45℃, which, in combination with our assay, allows measurement of rate constants up to 75℃. The assay was used to measure the enzyme properties of adapted adenylate kinases isolated from evolution experiments.

Decreasing the temperature ramp of the adaptation experiment allowed us to isolate novel thermostable mutants of adenylate kinase. Under strong selection [Couñago et al., 2006] requirements for keeping up with large temperature steps reduced the possibility of creating triple mutants since mutants with moderate improvements in fitness were not allowed to persist in the population. In the weak temperature ramp experiment double mutants were able to persist in the environment long enough to acquire tertiary mutations that benefited the fitness of the organism. However, because of the lack of precise temperature control of the experiment and reduced selection criteria the experiment suffered from being more stochastic than its predecessor. The increased fitness benefit of combining either the A193V or T179I mutations with Q16L can be explained by the importance placed on stability increases related to our choice of selection. In most circumstances mutations that increase activity but decrease stability are tolerated because natural genes tend to be buffered from small changes in stability. Since the *adk* system relies on thermal unfolding rather than decreased activity for selection, mutations that decrease stability will not provide a fitness advantage even if they process a modest increase in activity. This means that most mutations that were recovered that provided a fitness advantage for *G. stearothermophilus* had both an increase in activity and an increase in stability. Since the stability improvements had positive epistasis, having both of these mutations present was beneficial even in the absence of changes in activity. A different system is required to explore the effects of increases in enzyme function that are not directly tied

to increases in stability. Adaptation to antibiotic resistance provides a system where fitness is still strongly tied to an increase in function of our target enzyme, but is not limited to changes that provide stability as well as activity.

Chapter 4

Building a Growth Rate Function to Simulate the Effect of TetX2 Mutants on Fitness

4.1 Introduction

One of the fundamental questions of evolution concerns the physical basis for fitness differences between enzyme variants [Dykhuizen and Dean, 1990]. I am interested in how changes in the physicochemical parameters of enzymes affect the fitness of the organism as a whole. I developed a growth rate function for the effect of TetX2 on reducing growth rate inhibition from minocycline, using the physicochemical properties of the enzyme, to predict the effect of enzyme properties on fitness. The growth rate function allows us to bridge the link between the physical properties of enzymes and phenotypic changes. Using this model, I was able to predict the fitness function for TetX2 resulting from adaptive changes to enzyme properties, providing a

direct correlation between the cause (physicochemical changes to TetX2) and the effect (fitness). Mutants for TetX2 were obtained from both flask evolution and directed evolution studies performed by Kasia Walkiewicz [Walkiewicz et al., 2012]. We measured population histograms for the experimental evolution studies to reveal the success or failure of the TetX2 mutants *in vivo*. We used experimental evolution to validate the predictions made by my model and to provide additional insight that would allow us to extend and improve it. The growth rate function was derived empirically from studies about the mode of action of tetracycline antibiotics and their effects on growth inhibition [Thanassi et al., 1995, Argast and Beck, 1984, Argast and Beck, 1985, Chopra and Roberts, 2001, Chopra and Shales, 1981], and the mechanism of TetX2 action [Alekshun and Levy, 2007, Yang et al., 2004, Entsch and van Berkel, 1995]. Growth rates of TetX2 mutants were fit to kinetic measurements and used to predict values for parameters governing the growth rate function.

4.2 Serial Dilution Adaptation of *E. coli* Containing *tetX2* to Minocycline (MCN)

TetX2 is a monooxygenase-containing dehydrogenase that degrades tetracycline by hydroxylation of position C11a [Yang et al., 2004]. TetX2 has activity towards many first generation tetracyclines, such as oxytetracycline and chlortetracycline, and has modest ability to inactivate minocycline [Yang et al., 2004]. The MIC for minocycline for *E. coli* $\text{BW25113}_{\text{tetX2}}$ was measured at 16 μg/ml compared with a value of 256 μg/ml for tetracycline [Yang et al., 2004]. Adaptation of TetX2 to MCN was used to explore how incremental changes to the activity and affinity towards substrates via mutation can improve growth rate and, consequently, fitness. Adaptation of *E. coli*

containing *tetX2* to MCN was carried out in a serial dilution experiment over ten days by Kasia Walkiewicz [Walkiewicz et al., 2012]. The bacteria were transferred once a day to a fresh flask containing 50 ml LB with increasing concentration of MCN. The MCN concentrations for each of the 10 days were: 10, 16, 24, 36, 52, 80, 120, 180, 300, and 320 μg/ml. At the end of the experiment the TetX2 mutant $tetX2_{T280A}$ was fixed in the population. To isolate more mutants of TetX2 an error prone library was built and screened at 8 μM MCN in LB agar plates. The winner of the initial TetX2 experiment, $tetX2_{T280A}$, was recovered from the error prone library along with other novel mutants which match or increase the ability of TetX2 at degrading MCN. Seven mutants of TetX2 were identified using the error prone library; $tetX2_{T280A}$, $tetX2_{T280S}$, $tetX2_{F235Y}$, $tetX2_{K64R}$, $tetX2_{S236I}$, $tetX2_{N371I}$, and $tetX2_{N371T}$.

4.3 Kinetic Measurements of TetX2

TetX2 is a monooxygenase-containing a flavin cofactor that hydroxylates tetracyclines, requiring NADPH as an electron donor [Yang et al., 2004]. The active form of the enzyme is the reduced form $TetX2^{red}$, which degrades minocycline through a bisubstrate mechanism. $TetX2^{red}$ binds MCN in a binary complex and through a reaction involving molecular oxygen hydroxylates MCN, inactivating the antibiotic and oxidizing TetX2. $TetX2^{oxd}$ forms a binary complex with NADPH that recovers the active form $TetX2^{red}$. Although oxygen is required for the reaction, previous studies suggest we can assume that diffusion and binding of oxygen is fast enough that we can safely ignore its effects on kinetics [Entsch and van Berkel, 1995]. The reaction is therefore dependent on the Michaelis constant of TetX2 and MCN, $K_{M(MCN)}$, and of TetX2 and NADPH, $K_{M(NADPH)}$, as well as the reaction rate k_{cat} [Entsch and van

Berkel, 1995].

4.4 Growth Rate Measurements of TetX2 Mutants

Growth rates were measured by Christine Sun to measure the fitness advantage provided by the individual TetX2 mutants. Traditionally, fitness comparisons of antibiotic containing bacteria have been done by comparing the minimum inhibitory concentration (MIC) of drug required to prevent growth. MIC measurements can vary depending on criteria used to judge if bacteria are growing when the antibiotic is bacteriostatic or when there is resistance to the antibiotic. Furthermore, MIC measurements do not provide a precise measurement of fitness since measurements are qualitative (growth or no growth). A better measurement of the fitness advantage provided by TetX2 mutations is to measure growth rate dependence on MCN (growth rate curve). The growth rate curve provides a quantitative comparison of fitness between any two TetX2 mutants. Measurements of growth rates are also easily scalable with the use of a plate reader. Measurements were done using a plate reader over a period of 24 hours at different minocycline concentrations. Growth rates at exponential growth were calculated from each of the different MCN concentration growth curves using an R script provided by Rolf Lohaus. The R script measured the maximal growth rate using a linear fit of the exponential growth phase. Measurement of growth rates was complicated by the appearance of biphasic growth in bacterial populations that were undergoing strong inhibition. In cases of biphasic growth only the first linear growth regime was used, even if the second growth phase showed faster growth. The first growth phase was chosen because it represents the growth rate under antibiotic inhibition, while the causes of the appearance of the

Table 4.1: Steady state kinetic parameters for TetX2 and mutants

	$K_{M(MCN)}$ μM	$K_{M(NADPH)}$ μM	k_{cat} s^{-1}	$k_{cat}/K_{M(MCN)}$ μM^{-1} s^{-1}	$k_{cat}/K_{M(NADPH)}$ μM^{-1} s^{-1}	Δ Growth Rates at 32 μM MCN (%)[a]
WT	35 ± 1.9	75 ± 4.1	0.34 ± 0.01	0.010 ± 0.0006	0.004 ± 0.0003	100
T280A	18 ± 0.9	18 ± 1.1	0.43 ± 0.01	0.024 ± 0.0013	0.024 ± 0.0016	540
N371I	18 ± 1.9	64 ± 6.3	0.37 ± 0.02	0.020 ± 0.0024	0.006 ± 0.0006	530
N371T	24 ± 2.1	130 ± 11	0.40 ± 0.02	0.017 ± 0.0017	0.003 ± 0.0003	440
S326I	37 ± 2.8	73 ± 5.5	0.36 ± 0.01	0.010 ± 0.0008	0.005 ± 0.0004	340
F235Y	54 ± 6.1	99 ± 11	0.32 ± 0.06	0.006 ± 0.0013	0.003 ± 0.0007	320
K64R	36 ± 4.6	110 ± 15	0.32 ± 0.03	0.009 ± 0.0014	0.003 ± 0.0005	86
T280S	30 ± 3.4	100 ± 14	0.18 ± 0.01	0.006 ± 0.0008	0.002 ± 0.0003	65

[a]Changes in maximal growth rates of mutant alleles are relative to the wild-type tet(X2) growth rate set at 100%.

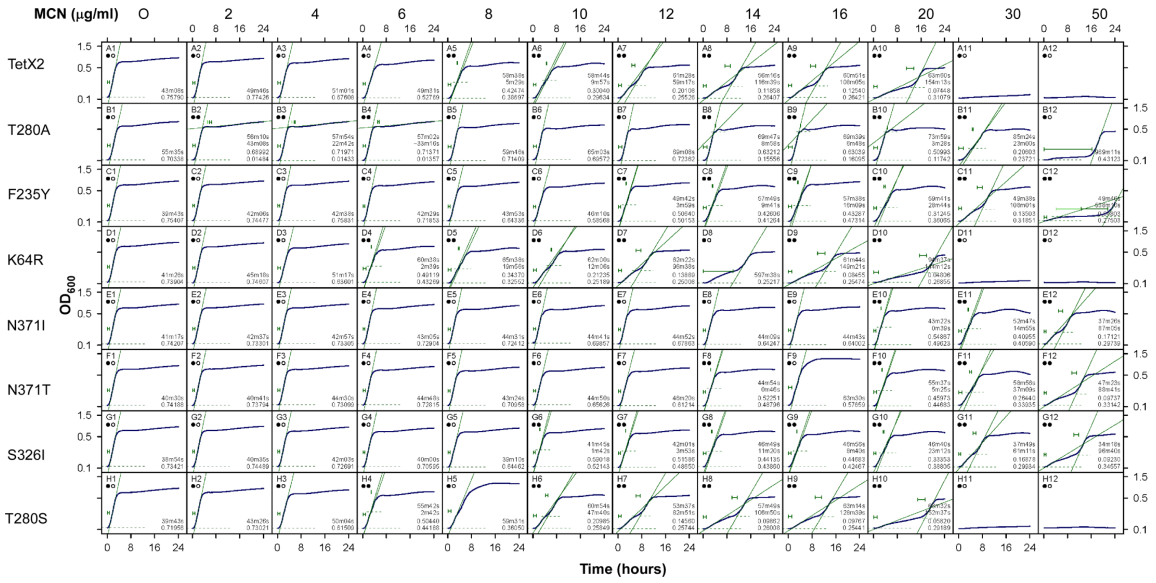


Figure 4.1: *E. coli* containing TetX2 were grown under minocycline inhibition and OD measurements were taken over 24 hours. Some of the growth curves displayed biphasic growth, with a second growth phase usually starting after 8 hours. The growth rates at exponential phase were estimated by calculating the maximum growth rate of the curve. In the cases where biphasic growth was observed only the first growth phase was used. [Walkiewicz et al., 2012]

second growth phase is likely due to secondary growth effects. Four possible causes for the appearance of a secondary growth phase are: the existence of a subpopulation of bacteria with increased growth rate that overtakes the initial population, the rise of new mutants, degradation of the antibiotic in the media over time due to oxidation, or transcriptional changes induced by the antibiotic. The plot of MCN vs maximal growth rates describes the fitness change provided by each mutant in the presence of MCN (growth rate curve) (figure 4.1).

Table 4.2: Relative Maximal Growth Rates[a] for TetX2 and Adaptive Mutants

MCN (μM)	*E. coli*$_{\mathrm{BW25113}}$	TetX2	TetX2$_{\mathrm{T280A}}$	TetX2$_{\mathrm{N371I}}$	TetX2$_{\mathrm{N371T}}$	TetX2$_{\mathrm{S326I}}$	TetX2$_{\mathrm{F235Y}}$	TetX2$_{\mathrm{K64R}}$	TetX2$_{\mathrm{T280S}}$
0	1.00 ± 0.01	1.00 ± 0.01	1.00 ± 0.01	1.00 ± 0.01	1.00 ± 0.01	1.00 ± 0.04	1.00 ± 0.03	1.00 ± 0.03	1.00 ± 0.02
4.1	0.59 ± 0.01	1.01 ± 0.01	0.97 ± 0.01	0.98 ± 0.01	0.98 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.97 ± 0.01	1.00 ± 0.01
8.1	0.39 ± 0.01	0.94 ± 0.03	0.99 ± 0.01	0.99 ± 0.01	0.98 ± 0.01	0.95 ± 0.01	0.96 ± 0.01	0.91 ± 0.05	0.92 ± 0.05
12.2	0.24 ± 0.01	0.73 ± 0.03	1.00 ± 0.01	0.99 ± 0.01	0.98 ± 0.01	0.93 ± 0.02	0.94 ± 0.01	0.69 ± 0.03	0.73 ± 0.03
16.2	0.09 ± 0.05	0.56 ± 0.03	1.00 ± 0.01	0.99 ± 0.01	0.96 ± 0.01	0.87 ± 0.02	0.85 ± 0.03	0.49 ± 0.03	0.52 ± 0.02
20.3	0.04 ± 0.01	0.41 ± 0.01	1.00 ± 0.02	0.97 ± 0.07	0.95 ± 0.07	0.79 ± 0.02	0.77 ± 0.02	0.29 ± 0.02	0.30 ± 0.02
24.3	ND	0.27 ± 0.01	1.00 ± 0.01	0.95 ± 0.04	0.87 ± 0.04	0.70 ± 0.02	0.67 ± 0.02	0.20 ± 0.01	0.21 ± 0.02
32.4	ND	0.17 ± 0.01	0.92 ± 0.02	0.89 ± 0.01	0.75 ± 0.01	0.58 ± 0.01	0.54 ± 0.01	0.12 ± 0.01	0.11 ± 0.02
40.6	ND	0.09 ± 0.01	0.78 ± 0.04	0.77 ± 0.01	0.62 ± 0.01	0.44 ± 0.01	0.40 ± 0.01	0.07 ± 0.01	0.06 ± 0.01
60.8	ND	ND	0.37 ± 0.05	0.55 ± 0.01	0.34 ± 0.01	0.20 ± 0.01	0.16 ± 0.01	ND	ND
101.4	ND	ND	0.07 ± 0.04	0.23 ± 0.02	0.12 ± 0.02	0.10 ± 0.02	0.03 ± 0.02	ND	ND

[a] Growth rates in the table are normalized for each strain tested to the growth rate determined at 0μM MCN.

ND = growth not detected

[Walkiewicz et al., 2012]

4.5 Mathematical Model for Mass Transfer of Minocycline

To investigate the link between kinetic measurements and growth rates we have developed a growth rate function describing TetX2's contribution to minocycline resistance. MCN inhibits cell growth by reversibly binding to and inactivating ribosomes [Chopra and Roberts, 2001]. TetX2 does not protect the ribosome from the effects of MCN, TetX2 instead provides a fitness benefit by reducing the steady state concentration of MCN in the cell [Whittle et al., 2001]. Two factors need to be considered to understand the effect of TetX2 at reducing MCN concentrations: the diffusion of MCN from the bulk media into the cytoplasm, and the actual inactivation of intracellular MCN by TetX2. The balance of the two components, influx of TetX2 by diffusion and removal of MCN through TetX2 inactivation, provides a mass transfer equation to estimate the concentration of MCN in the cell [Thanassi et al., 1995]. We can model the ability of TetX2 to reduce the cytosolic concentration of MCN (MCN_C) by calculating the intracellular concentration of TetX2 and its associated kinetics. When the MCNc of two mutants is equal, regardless of what the extracellular MCN concentration is for each of the mutants, the two mutants will have the same growth rate because the enzyme does not affect MCN binding to the ribosome. An equation that relates MCN concentration and growth rates by measuring the effect of MCN binding to the ribosome on inhibiting growth rates was used to complete the growth rate function.

Diffusion of minocycline through the outer and inner cell membranes can be modeled using Fick's law of diffusion. Tetracyclines typically cross the outer membrane through porins after binding Mg^{2+} (figure 4.3). The difference in the ionic potential

between the periplasm and the media, called the Donnan potential, drives MCN into the periplasm. Tetracycline concentration in the periplasm is estimated to be twice that of the bulk media at equilibrium [Thanassi et al., 1995]. We have used this same value for MCN, but it is likely that MCN diffusion will be slightly different due to it's slightly different pKa's [Thanassi et al., 1995, Parolo et al., 2010]. Diffusion through the inner membrane is dependent on the concentration of the uncharged species of MCN. The pH difference between the periplasm and the cytoplasm means the concentration of uncharged MCN is higher in the periplasm than the cytoplasm. The effective concentration in the cytoplasm is again 1/2 as much as the actual concentration in the periplasm. Because the concentration gradient across each of the membranes is not 1:1 but is instead 2:1, MCN accumulates inside the cell to a concentration of 4:1 compared to the bulk media [Thanassi et al., 1995]. In the absence of any other force, minocycline concentration in the cytoplasm (MCN_C) would be 4 times the concentration in the media (MCN_M) at steady state. However, degradation of MCN works against diffusion to decrease the steady state concentration.

The activity of TetX2 is calculated by using Michaelis-Menten steady state kinetics. The TetX2 reaction involves a two state reaction with a "ping pong" or bisubstrate mechanism. Studies of the similar reaction mechanism of *p*-hydroxybenzoate hydroxylate (PHBH) [Entsch and van Berkel, 1995], another flavoprotein, show the binding rate of O_2 is fast enough that it is not a rate limiting step. Oxygen levels remain relatively constant during the experiment, further allowing us to focus instead on the effects of NADPH and MCN on reaction rates. The reaction begins with MCN binding to the reduced form of the free enzyme to form the enzyme MCN complex with equilibrium constant $K_{M(MCN)}$ (figure 4.2). The complex then reacts rapidly with O_2 allowing us to simplify the hydroxylation reaction to a pseudo first

order reaction with reaction rate k_1. Products of the reaction are the oxidized form of the enzyme and the dissociation of the hydroxylated MCN (MCN*) (Reaction R 4.1). Hydroxylated MCN is unstable at the pH of the cytoplasm and quickly degrades non-enzymatically [Yang et al., 2004]. The enzyme has to be returned to the reduced form from it's oxidized state by the reductant NADPH to become active again. The binary complex of the oxidized enzyme and NADPH binds with a Michaelis constant defined as $\mathrm{K_{M(NADPH)}}$. The complex then reacts with a reaction rate k_2 to form the reduced enzyme and $NADP^+$ (reaction R 4.2). Lineweaver-Burk plots of the substrates NADPH and MCN reveal that both compounds compete for the same binding site, or overlapping sites. Parallel lines in the Lineweave-Burk plots suggest that only binary complexes are occurring, which is consistent with the "ping-pong" mechanism.

$$\mathrm{TetX2}^{red} + \mathrm{MCN} \overset{K_{M(MCN)}}{\rightleftharpoons} \mathrm{TetX2}^{red} : \mathrm{MCN} \xrightarrow[+\mathrm{O_2}]{k_1} \mathrm{TetX2}^{ox} + \mathrm{MCN*} \qquad \text{(R 4.1)}$$

$$\mathrm{TetX2}^{ox} + \mathrm{NADPH} \overset{K_{M(NADPH)}}{\rightleftharpoons} \mathrm{TetX2}^{ox} : \mathrm{NADPH} \xrightarrow{k_2} \mathrm{TetX2}^{red} + \mathrm{NADP} \qquad \text{(R 4.2)}$$

The Michaelis-Menten equation for a bisubstrate reaction is

$$v_o = \frac{\frac{v_{max}}{1+\frac{K_{M(B)}}{[B]}}[A]}{\frac{K_{M(A)}}{1+\frac{K_{M(B)}}{[B]}} + [A]} \qquad (4.1)$$

Where MCN is substrate A and NADPH is substrate B [Cleland, 1963]. The derivation of the bisubstrate rate equation appears in section 2.8.2.

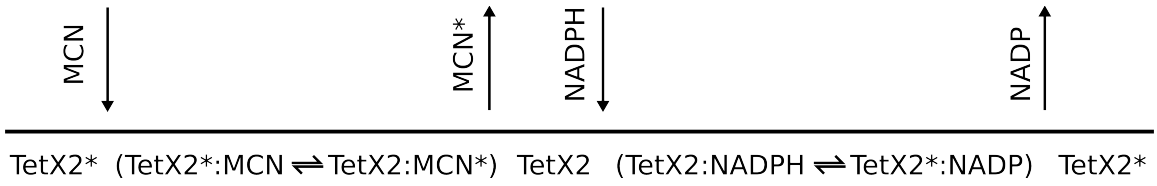


Figure 4.2: TetX2 degradation of minocycline occurs through a "ping-pong" or binary mechanism. The reduced form of TetX2 ($TetX2^{red}$ or TetX2*) reacts with minocycline to form the hydroxylated product MCN*. NADPH reacts with TetX2 to produce the activated form of the enzyme and NADP.

4.6 A physiological model for TetX2 mediated inactivation of MCN

We developed a growth rate function based on experimental data to quantitatively describe the success and failure of each adaptive mutation over a range of MCN concentrations from their *in vitro* properties (figure 4.3). As shown in figure 4.8, growth rates for E. coli BW25113 carrying variants of *tetX2* could be predicted accurately from experimentally derived kinetics and protein expression levels. Inhibition of bacterial growth rates by cytosolic MCN (MCN_C) was determined by fitting the growth rate dependence of E. coli to MCN with equation 2.25.

Growth rate inhibition by MCN_C is predicted to be dependent on the ratio of uninhibited ribosomes to total ribosomes. We can assume that the growth rate of bacteria, at exponential phase, will roughly increase linearly with free ribosome concentration as a proportion of total protein [Ecker and Schaechter, 1963]. Under MCN inhibition the total fraction of available ribosomes will be equal to 1 minus the ratio of inhibited, or substrate bound, ribosomes. The Hill function (equation 2.25) is a general biochemical function for substrate binding that can be used to calculate the fraction of uninhibited ribosomes and the fraction of ribosomes that are being

inhibited by binding MCN. The Hill function was developed originally to provide an empirical measurement of binding of oxygen to hemoglobin [Weiss, 1997]. The function has been extended to apply to more general measurements of ligand-receptor interactions. In our case we can use the Hill equation to provide an empirical measurement of the binding of MCN to the ribosome, as well as to the functional effect of subsequent binding events on reducing the growth rate of bacteria.

$$GR = 1 - \frac{[\mathrm{MCN}_C]^B}{A + [\mathrm{MCN}_C]^B} \tag{2.25}$$

At steady state the concentration of NADPH and MCN will be constant, providing a constant degradation rate of MCN to counteract diffusion [Thanassi et al., 1995]. In *E. coli* BW25113, endogenous MCN export was assumed to be negligible compared to degradation by TetX2 such that at steady state the rate of diffusion of minocycline equals the degradation rate of MCN by TetX2 (equation 2.22).

$$\frac{V_{max}^{\mathrm{obs}}[\mathrm{TetX2}][\mathrm{MCN}_C]}{K_M^{\mathrm{obs}} + [\mathrm{MCN}_C]} = D_I \left([\mathrm{MCN}_P] - \frac{1}{2}[\mathrm{MCN}_C] \right) \tag{2.22}$$

The bisubstrate kinetics equation (equation 2.17) and Ficks law (equation 2.10) were used to calculate the steady state cytosolic MCN ($\mathrm{MCN_C}$) concentration from the concentration of MCN in the media (equation 2.24). Derivation of the steady state equation is detailed in section 2.8.1.

$$\begin{aligned}[\mathrm{MCN}_C] = 2 \left([\mathrm{MCN}_O] - 0.25K_M^{\mathrm{obs}} - \frac{V_{max}^{\mathrm{obs}}}{D}[TetX2] \right) \\ - 2\sqrt{\left([\mathrm{MCN}_C] - 0.25K_M^{\mathrm{obs}} - \frac{V_{max}^{\mathrm{obs}}}{D}[TetX2] \right)^2 + K_M^{\mathrm{obs}}[\mathrm{MCN}_C]}\end{aligned} \tag{2.24}$$

where

$$V_{max}^{\text{obs}} = \frac{V_{max}[\text{NADPH}]}{K_M^{\text{NADPH}} + [\text{NADPH}]} \tag{2.18}$$

$$K_M^{\text{obs}} = \frac{K_M^{\text{MCN}}[\text{NADPH}]}{K_M^{\text{NADPH}} + [\text{NADPH}]} \tag{2.19}$$

We assumed that the concentration of NADPH in the cell is constant because cellular metabolism constantly regenerates the NADPH pool, and was not likely to be altered by addition of MCN. The NADPH concentration is therefore estimated to be 150 μM [Andersen and von Meyenburg, 1977]. The concentration of TetX2 in the cell will also determine the degradation rate since $\text{v}_{\text{max}} = \text{k}_{\text{cat}} \times [\text{TetX2}]$. While we do not know what the concentration of TetX2 is inside the cell we are able to measure the relative concentrations of each mutant protein. Changes in both protein expression levels and stability between each of the different mutants may alter the amount of active protein that is present in the cell. Unfortunately, we were unable to measure the temperature dependent stability of the protein because the protein unfolds irreversibly and the multiple transitions were observed in CD signal under thermal unfolding [Walkiewicz et al., 2012]. However, we were able to measure the relative protein concentrations of each of the mutants (table 4.3). The protein concentrations were calculated by measuring the enzyme activity in cell extracts towards MCN and normalizing for total protein, measured using a Bradford assay. These protein concentrations give us a measure of the relative activity for each of the mutants with a given k_{cat}. The caveat of this measurement is that it is dependent on accurate measurements of k_{cat}. The relative rates of diffusion and enzyme catalysis in the growth rate function are inversely proportionate since these two values are always present in the form of $\frac{V_{max}}{D}$ in equation 2.24. The diffusion parameter D is therefore more accurately described as the ratio of diffusion rate to protein concentration. This value is also modified

Table 4.3: Normalized Concentration and Total Activity for TetX2 Mutants From Cell Extracts

	Concentration*	Activity (μMs^{-1})
TETX2	1.00	0.16
T280A	0.77	0.19
F235Y	2.00	0.22
K64R	1.14	0.14
N371I	1.36	0.27
N371T	1.44	0.25
S326I	1.56	0.24
T280S	2.03	0.16

*Concentrations were calculated using the measured k_{cat} and normalized to TetX2 wild type

by the fact that the diffusion rate scales with the surface area of the cell, while the protein concentration instead is inversely proportional to the volume of the cell. D is therefore a lumped parameter that accounts for the diffusion constant, normalized protein concentration, and cell volume that is assumed to be constant among the mutants. Therefore, the unknown parameters of the system are the combined diffusion rate, D, and the Hill function parameters, A and B, which describe the interaction between MCN_C and growth rate. At higher concentrations of MCN, inhibition of cell growth is a consequence of the limited ability of TetX2 or a particular variant to inactivate MCN to tolerable levels.

The Hill function parameters, A and B, were determined by fitting the Hill function to the mutant's experimental growth rates plotted against MCN_C. Each MCN_C point was calculated from the corresponding MCN_O point for each mutant by using

equation 2.25 and the measured values of $K_{M(MCN)}$, $K_{M(NADPH)}$, and k_{cat}(table 4.1). The parameter A corresponds to the apparent dissociation constant K_D for minocycline binding to the ribosome. K_D in the context of the model represents interactions between MCN and the ribosome but is also affected by other MCN interactions within the cytosol that can compete with MCN binding to the ribosome. Despite this caveat, the value of A was estimated to be about 26 μM, which is in good agreement with the MCN concentration at which adaptive mutants were isolated (32 μM). The parameter B is the Hill coefficient that represents how the effect, on the growth rate at exponential phase, of increasing MCN concentration deviates from a linear response. Parameters for A, B, and the combined diffusion rate D were fitted using a least squares method to find the best fit for estimating growth rates. An optimization function was used to find the parameters that minimized the sum of squared (SS) difference between the measured and predicted growth rate measurements.

$$\text{SS} = \sum (\text{Measured Growth Rate} - \text{Predicted Growth Rate})^2 \tag{4.2}$$

A mathematical transformation was used to linearize the curve to improve the response of the model to edge effects [Zarr, 1999].

$$(\text{Growth Rate})_{transformed} = arcsine(\text{Growth Rate}^2) \tag{4.3}$$

The arcsin transformation stretches the edges of the model, which improves the model's ability to fit the edges of the curve. Growth rate curves were transformed back after the fit was finished, so that there is no effect on the final shape of the curve. Growth rates that were higher than the normalized initial growth rate (normalized to 1.0 at 0 MCN) were normalized to be equal to 1.0, since that is defined

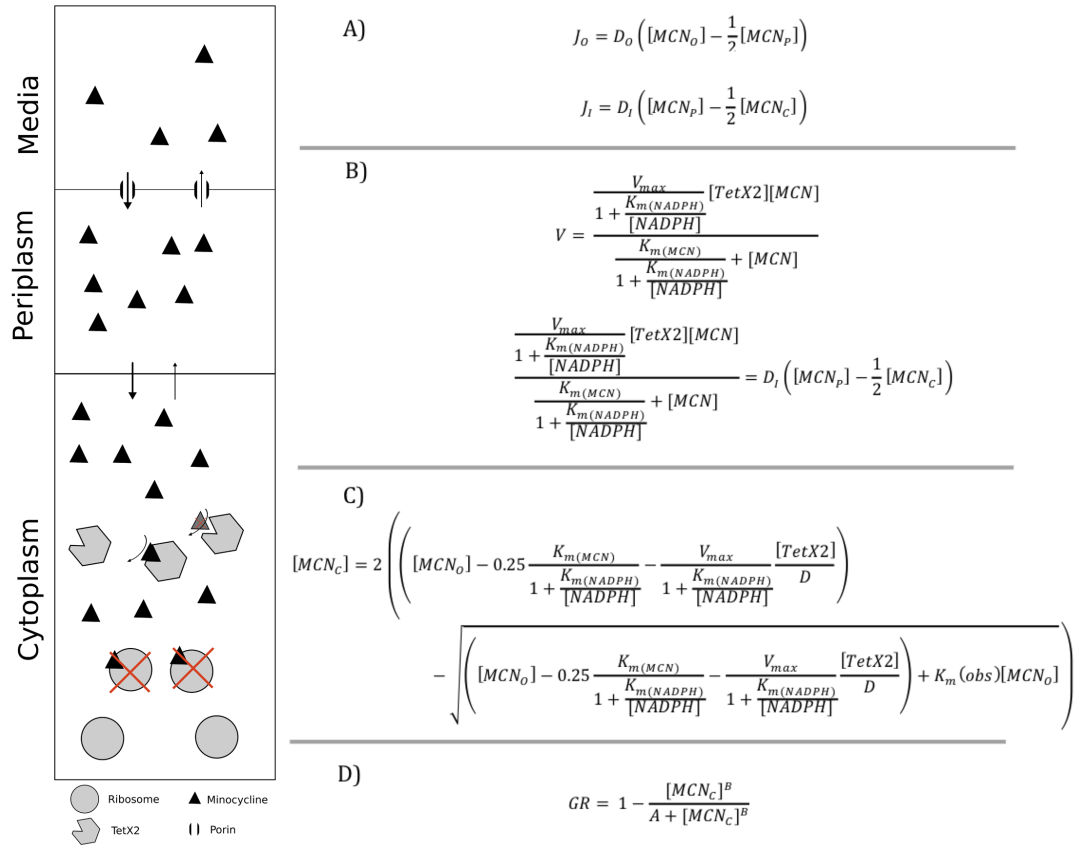
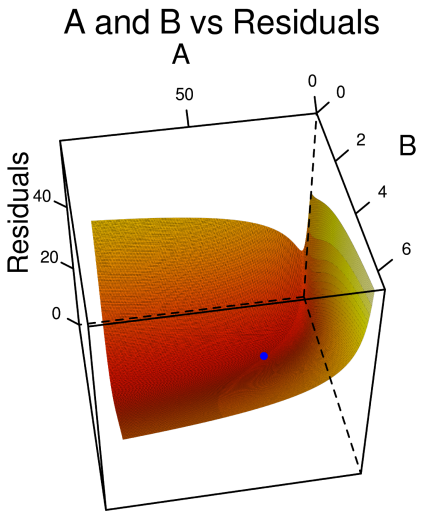


Figure 4.3: A) MCN flux through the cell membrane is controlled by passive diffusion. The difference in Mg_2^+ concentration and pH across the outer and inner membranes respectively affects the concentration of species that can diffuse across the membrane. We can estimate the apparent concentration of MCN in the periplasm to be $\frac{1}{2}$ that of the concentration in the media and further, that the apparent concentration of MCN in the cytoplasm is once again $\frac{1}{2}$ that of the periplasm [Thanassi et al., 1995]. B) The rate of degradation by TetX2 is calculated using the Michaelis-Menten equation for a bisubstrate reaction. At steady state the rate of diffusion is equal to the degradation rate by TetX2. C) Solving the quadratic equation for the concentration of MCN inside the cell results in a function that depends on the combined diffusion constant D and kinetic parameters. D) The effect of growth rate inhibition by MCN is calculated from the concentration of MCN inside the cell and the variables A and B by using the Hill equation where growth rate is proportional to the uninhibited ribosome concentration.

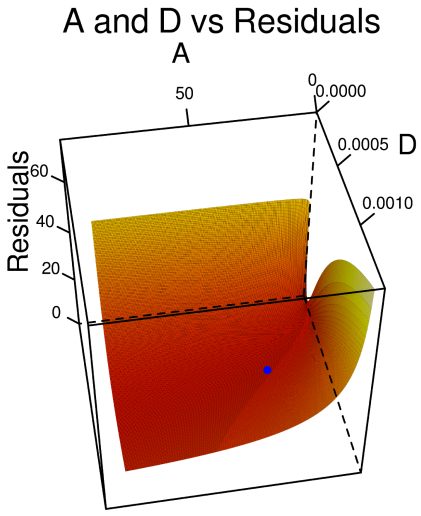
as the maximal normalized growth rate of the bacteria. The mathematical fits were done using the R programming language and the optim function with standard parameters [Ihaka and Gentleman, 1996]. The function optim finds the best value of A, B, and D that minimizes the value of the total sum of squares (SS). The R function optim uses the Nealder-Mead algorithm for optimization. The initial guesses for the model parameters were 20, 6, and 0.01. The predicted parameter values were: $A = 25.8$, $B = 2.5$, $D = 6.7 \times 10^{-4}$. The A parameter and the accumulation of MCN inside the cytoplasm are correlated since decreasing A can roughly compensate for an increase in MCN accumulation. We can observe the correlation between the parameters A and D by calculating the sum of squared differences for the surface on which the mathematical fit occurs (figure 4.4). The surface for A and D is much flatter than the surface for A and B or B and D. The solution occurs at the intersection of the three two dimensional surfaces. Bacteria with significant changes to membrane chemistry or transporters could have a different effect on the interactions of the three parameters.

4.7 Measuring Population Distributions Using High Throughput Sequencing

In order to test if any of the other TetX2 mutants can have success in rising in frequency in the flask population, ten more sets of serial dilution adaptation experiments were carried out by Kasia Walkiewicz [Walkiewicz et al., 2012]. These experiments were carried out over 3 days with MCN concentrations of 10, 16, and 24 μg/ml. Walkiewicz et a., showed that kinetic measurements and growth rate measurements suggested that the mutants $tetX2_{T280A}$, $tetX2_{N371I}$, and $tetX2_{N371T}$ provide similar

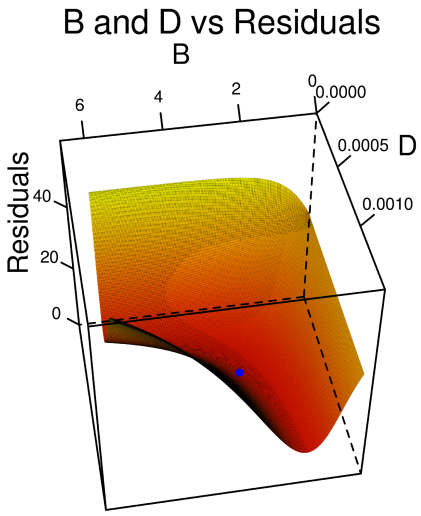

A and B vs Residuals
A
50
0
0
2
B
4
6
Residuals
40
20
0

(a)


A and D vs Residuals
A
50
0
0.0000
0.0005
D
0.0010
Residuals
60
40
20
0

(b)

Figure 4.4



(c)

Figure 4.4: Surfaces showing the sum of squared difference values between experimental and model growth rate curves for different values of A, B, D. Each of the surfaces reveals a furrow for solutions in 2 dimensions; a) A and B, b) A and D, and c) B and D. The fit is able to converge at the point where the three surface intersect. The point represents the value at each surface where the minimum of the three dimensional surface of A, B, and D exists.

benefits to fitness (table 4.1, figure 4.1) [Walkiewicz et al., 2012]. Assuming that there is no major difference in mutation rate for each of the three mutations, all three mutations should provide sufficient benefit to appear in the population. We carried out a deep sequencing experiment to build a histogram of *tetX2* mutants in each of the three experiments (figure 4.7). To measure population dynamics we used a sequencing method that exploits the large number of short reads that are produced by next generation sequencing technologies, specifically Illumina sequencing (FREQ-SEQ) [Chubiz et al., 2012].

Illumina short read sequencing technology was first developed for genome sequencing but has been adapted by Chubiz et al., to monitor allelic frequencies in mixed populations [Bentley et al., 2008, Chubiz et al., 2012]. The genome is first digested into smaller fragments, each fragment is barcoded and the mixture is sequenced using a multiplexed method [Bentley et al., 2008]. Illumina forward and reverse primers are added to each of the fragments before sequencing. The forward and reverse primer regions are complementary to oligonucleotides that are immobilized on the Illumina flowcell surface. Fragments are amplified on the plate such that each of the fragments occupies a distinct spot on the plate. Because each section of the flowcell surface contains numerous fragments sequencing is multiplexed through reading multiple spots at each sequencing cycle [Bentley et al., 2008]. The entire sequence is stitched together when the overlapping fragments from each segment are used to order the individual sequencing reads computationally. The strength of this system is the ability to read multiple sequences simultaneously and provide a high enough read density to be able to piece together a genome by ensuring that each fragment of DNA is read multiple times. The coverage for an entire *de novo* bacterial genome is usually onf the order of 20 to 40 times [Bentley et al., 2008].

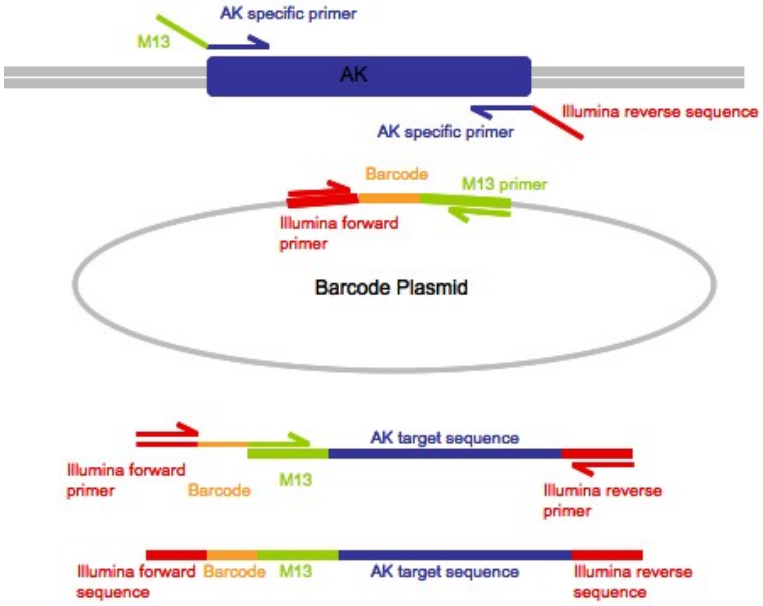


Figure 4.5: First the region containing the target gene is first amplified from the chromosome. The region containing the target nucleotide is amplified off the gene with flanking regions containing and M13 sequence in the 5' end and the Illumina reverse sequence on the 3' end. The barcode sequence is amplified off a plasmid with the Illumina forward sequence on the 5' end and the complementary reverse M13 sequence on the 3' end. The barcode and the target oligonucleotide are amplified together with an Illumina forward and reverse primer to form the final oligonucleotide for sequencing. Adapted from diagrams obtained from Lon Chubiz [Chubiz et al., 2012].

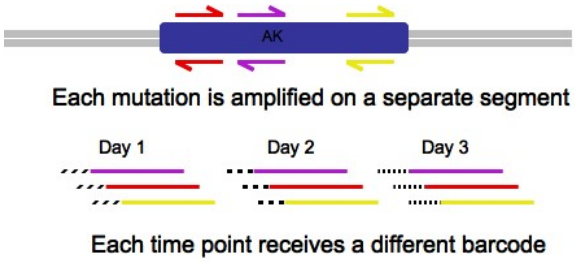


Figure 4.6: Each of the 5 target nucleotides was amplified on a separate segment of DNA for each day for a total of 81 fragments. The fragments were labeled with unique 6 nucleotide barcodes that identify the day of the experiment. The primer for each segment identifies the target nucleotide site, and the barcode identifies the day of the experiment when the sequencing results are analyzed. Adapted from diagrams obtained from Lon Chubiz [Chubiz et al., 2012].

FREQ-SEQ exploits the large number of possible reads to instead repeatedly sequence near-homogenous sequences to measure ratios of single nucleotide variation [Chubiz et al., 2012]. Instead of having an entire genome broken into random segments specifc oligonucleotides that includes the target region for sequencing are used to monitor SNP and INDEL frequency. The oligonucleotide also included a 6 nucleotide identifier sequence, or barcode, to identify the sample (a time point in the case of our experiment, for example, day 1). The barcodes were generously donated by the Marx Laboratory at Harvard [Chubiz et al., 2012]. Oligonucleotides containing each of the target mutation positions were amplified and then barcoded with the corresponding time point in separate reactions. 31 barcodes were used to identify each day of the ten populations. There were a total of 3 time points each for days 1, 2, and 3, and an initial starting population at day 0. Each mutation site, *a808* (T280), *a1082* (N371), *t674* (F235), *g947* (S326), and *a161* (K64), required a different oligonucleotide. In total there were $31 \times 5 = 155$ separate oligonucleotides for sequencing. Because of the small number of individual DNA fragments compared to a standard Illumina genome sequencing project coverage of each nucleotide increased to about 10^3 to 10^6 per nucleotide, giving us an extremely accurate measurement of the ratio of nucleotides at each position. Sources of potential error include polymerase error during amplification and sequencing chemistry. The distribution of A, T, G, and C bases of positions that did not provide beneficial mutations was used to estimate the confidence interval for identifying population distributions. We estimate that the coverage for each oligonucleotide was on the order of 10^3 to 10^4 individual reads [Erlich et al., 2009]. Measurement of the distribution at each position was calculated to be within 0.5% error [Walkiewicz et al., 2012].

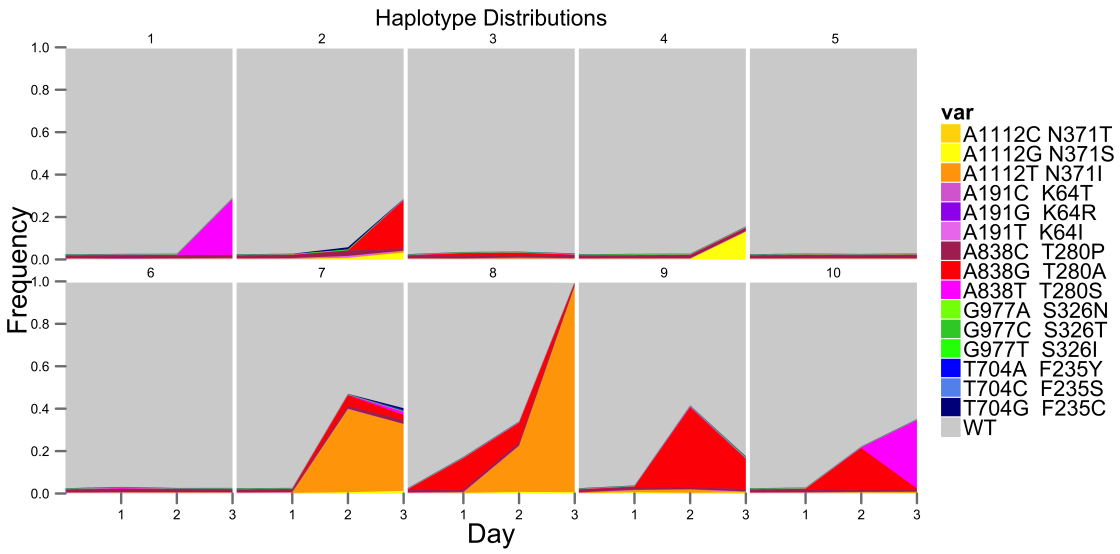


Figure 4.7: Deep sequencing was used to monitor the frequencies of adaptive *tetX2* variants (identified by error-prone mutagenesis) in ten individual populations of *E. coli* $\mathrm{BW25113}_{\mathrm{tetX2}}$ evolved for three days to increasing MCN concentrations (10, 16 and 24 μg/ml). The plot represents the relative frequencies of each of the 4 possible single nucleotide replacements fore each of the 5 nucleotide positions where mutations were found, A1082, A161, A808, G974, and T674. As predicted by the model, mutation $tetX2_{T280A}$ is the most successful and is present in five (out of 10) populations, while $tetX2_{N371I}$ is found in two populations and reaches fixation in one of them (Population 8). An unexpected neutral allele $tetX2_{T280S}$ is present in two populations. The remaining mutants do not appear or remain at very low frequencies. The possibility of mutations outside of *tetX2* providing fitness benefits results can account for variability of evolutionary outcomes.

4.8 Conclusion

The growth rate function was able to accurately predict the fitness effects of TetX2 adaptive mutants that were recovered using directed evolution. We were able to accurately model the growth rates for the 8 *tetX2* variants isolated from random mutagenesis using a mathematical description of the fitness effects from physical properties of the enzyme (figure 4.8). Growth rates of $tetX2_{F235Y}$, $tetX2_{S326I}$, and $tetX2_{T280S}$ could be predicted reliably only if the intracellular protein concentration was increased considerately from that of wild type TetX2 (table 4.3). Measurements of population dynamics from replicate populations of *E. coli* containing *tetX2* provide evidence that other adaptive mutants besides $tetX2_{T280A}$ can have success in the population. $TetX2_{N371I}$ was equally successful compared with $tetX2_{T280A}$ in both the growth rates measurements and the population distribution measurements. However, the appearance of mutations such as $tetX2_{T280S}$, which does not provide a fitness benefit when compared to WT *tetX2*, reveals our fitness model is still incomplete. The discrepancy between the predicted fitness and the apparent fitness of *E. coli* can be explained in two ways; either there is an unknown fitness effect that is provided by the mutation TetX2 T280S that is not caused by a change in kinetic parameters or protein concentration, or there is a mutation that was not present in *tetX2* that allowed those populations to rise in frequency, for example, a mutation in the promoter region of the enzyme. The probability for mutations outside of *tetX2* increasing fitness is high since the mechanisms for antibiotic resistance are complex. There is, however, no apparent reason for why $tetX2_{T280S}$ would show positive epistasis with these mutations while other *tetX2* mutants do not. An increase in the frequency of the TetX2 T280S mutation might account for the higher appearance of the mutant, but the mutation rate would have to be several orders of magnitudes bigger than

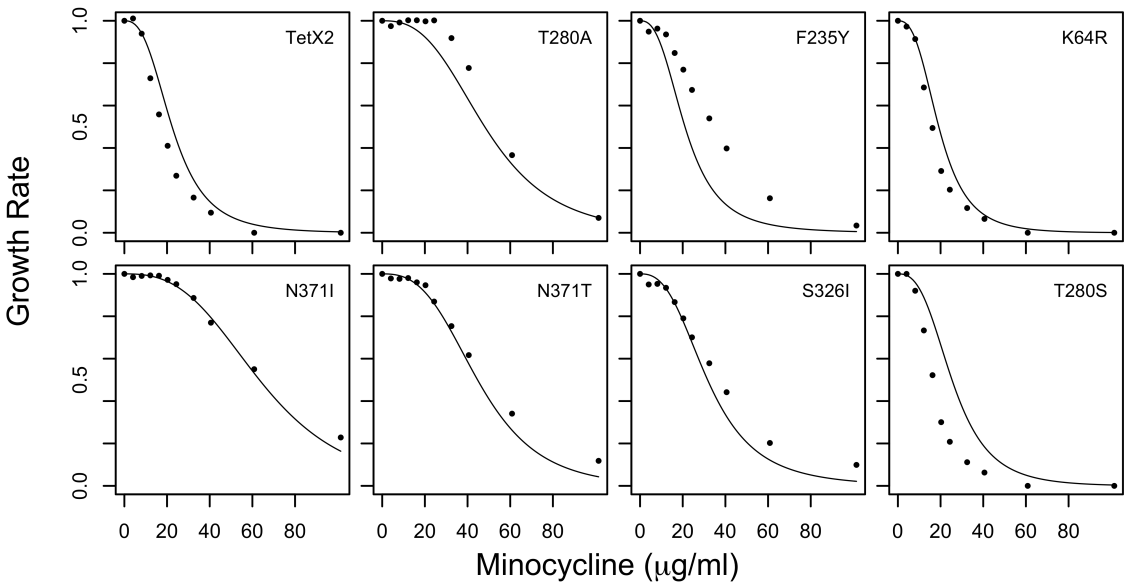


Figure 4.8: A growth rate function was built to predict the effect of TetX2 on reducing the inhibition of growth rates by MCN. By transforming equation 2.25 with equation 2.24 we were able to accurately fit the measured growth rate using *in vitro* kinetic parameters and estimates of protein concentration.

the mutation rate of the other mutants. Comparison of whole genome sequencing of the unadapted and adapted strains would be the best method to determine what mutations outside of *tetX2* are affecting the fitness of the organism.

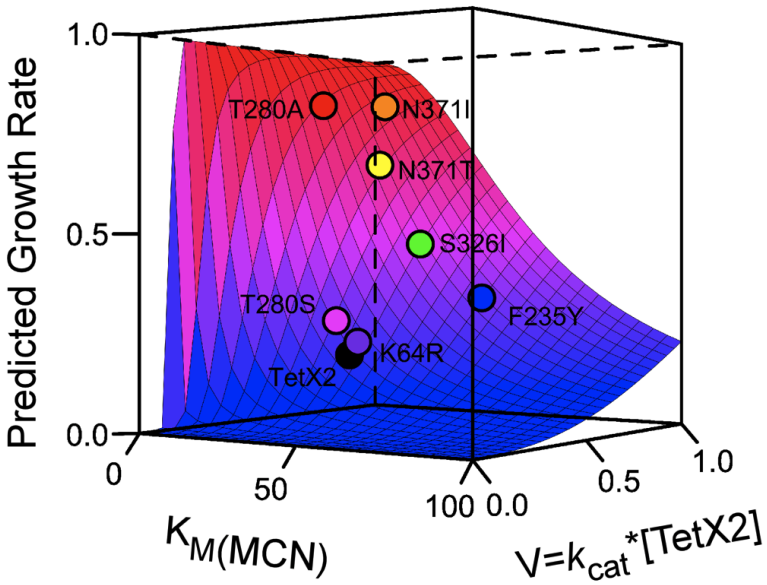


Figure 4.9: The physicochemical fitness landscape shows how changes in $K_{M(MCN)}$ and enzyme activity can affect growth rates at 32 μM MCN. Predicted growth rates from measured kinetic parameters for TetX2 (black), $TetX2_{T280A}$ (red), $TetX2_{N371I}$ (orange), $TetX2_{N371T}$ (yellow), $TetX2_{S326I}$ (green), $TetX2_{F235Y}$ (blue), $TetX2_{K64R}$ (purple) and $TetX2_{T280S}$ (magenta) are highlighted on the surface of the fitness landscape. [Walkiewicz et al., 2012]

Chapter 5

Using the growth rate function to Provide Insights into Adaptation

5.1 Introduction

We used the growth rate function to determine how the biophysical properties of TetX2 contribute directly to fitness during selection to minocycline (MCN). The parameters for the growth rate function that were calculated through the initial fit for A, B, and D were used to back-calculate kinetic parameters as a proof of principle showing that the modeling process is reversible. We demonstrate that the growth rate function can be used to estimate kinetic parameters of unknown mutants without requiring extensive protein purification and kinetics assays. *In vitro* characterization of enzymes remains the most accurate method for estimation of kinetic parameters. The growth rate function allows us to rapidly estimate kinetic parameters. In practice a small number of enzymes covering a range of kinetic measurements would be required to populate the initial model and then with the fitted parameters we could charac-

terize new variants without additional *in vitro* studies. We also investigate the error of the growth rate function on predicting the kinetic parameters using a statistical method. The relationship of fitness to each of the three main kinetics parameters, $K_{M(NADPH)}$, $K_{M(MCN)}$, and V_{max} is also predicted using the growth rate function.

5.2 Calculating Kinetic Parameters from Growth Rates

To validate the accuracy of the mathematical fit, the model parameters were used to back calculate $K_{M(MCN)}$ and V_{max} for each of the mutants from the curve of growth rates at exponential phase (growth rate curve). We focused on predicting $K_{M(MCN)}$ and V_{max} because these two parameters were shown to be more important in determining fitness values than $K_{M(NADPH)}$. Analysis of the effect of changing the various parameters indicates the relative importance of each of them on adaptation (table 5.1). Increasing the V_{max} or decreasing the $K_{M(MCN)}$ causes a significant improvement in the growth rate curve. At 32 μg/ml MCN, for TetX2 wild type, a 10% increase in k_{cat} causes a 23% increase in growth rate at exponential phase. Under the same conditions a 10% decrease in $K_{M(MCN)}$ also increases the exponential phase growth rate by 17%, but the same 10% increase in $K_{M(NADPH)}$ only increases the growth rate by 1% (table 5.1). $K_{M(MCN)}$ has a significant effect in increasing the length of the plateau of the growth curve before which the growth rate starts to decrease significantly (figure 5.3A). Decreasing $K_{M(NADPH)}$ will also cause a similar effect to decreasing $K_{M(MCN)}$, however, to a lesser extent (figure 5.3C) since the *in vivo* concentration of NADPH is well in excess of the range of MCN concentrations we used. Conversely V_{max} affects the slope of the growth rate curve decay (figure 5.3B). The V_{max}, or the catalytic rate,

can be increased either by increasing the overall reaction rate k_{cat} or by increasing the amount of active protein in the cell. Increasing either the stability of the enzyme or the amount of protein expression will increase V_{max} by causing an increase in the steady state concentration of TetX2.

The growth rate curves for each of the mutants was fit independently using a two-step approach. We began by using an arcsin transformation on the growth rates to reduce edge effects. Both the measured growth rate curves and the predicted curves were transformed using the same equation. The growth rate curves were transformed back after the mathematical fit predictions were finished. Initial tests showed that the model predictions depended on the initial parameter guess given to the Nelder-Mead algorithm. Further investigation suggested that this was because the solution landscape for many of the mutants when varying V_{max} and $K_{M(MCN)}$ contained a furrow that corresponded to a specific value of the ratio $V_{max}/K_{M(MCN)}$ (figure 5.3A). To improve our ability to find a consistent solution for the fits of kinetic parameters, we introduced the value of activity = $V_{max}/K_{M(MCN)}$ as a second fitting parameter. An initial fit was then done using a wide range of starting conditions for $K_{M(MCN)}$ and activity ($V_{max}/K_{M(MCN)}$). 50 values for each of the two variables were used with $K_{M(MCN)}$ ranging from 10 to 100, and $V_{max}/K_{M(MCN)}$ ranging from 0.001 to 0.02. The initial conditions were chosen to exhaustively cover the biologically relevant values for the two parameters by comparison to the measured kinetics. A total number of 2500 initial conditions were tested for each mutant. The result was a set of solutions that share a similar $V_{max}/K_{M(MCN)}$ but could vary in absolute values of V_{max} and $K_{M(MCN)}$ (figure 5.1A). As shown in figure 5.1, the shape of growth rates curves generated from an ensemble of equal $V_{max}/K_{M(MCN)}$ ratios of varying absolute magnitudes of V_{max} and $K_{M(MCN)}$ produce different growth rate responses that can be used as the second

Table 5.1: Effect of Improving Kinetic Parameters on Growth Rates

Effect of Decreasing $K_{M(NADPH)}$ by 10% on Growth Rate

	MCN (μM)		
	4	32	64
TetX2 WT with $K_{M(NADPH)}$ = 75	0.99	0.26	0.04
TetX2 WT with $K_{M(NADPH)}$ = 67.5	0.99	0.26	0.04
$TetX2_{T280A}$ with $K_{M(NADPH)}$ = 18.3	1.00	0.77	0.33
$TetX2_{T280A}$ with $K_{M(NADPH)}$ = 16.5	1.00	0.77	0.33

Growth rates were calculated using the growth rate function

Effect of Decreasing $K_{M(MCN)}$ by 10% on Growth Rate

	MCN (μM)		
	4	32	64
TetX2 WT with $K_{M(MCN)}$ = 35.4	0.99	0.26	0.04
TetX2 WT with $K_{M(MCN)}$ = 31.9	0.99	0.30	0.05
$TetX2_{T280A}$ with $K_{M(MCN)}$ = 18.0	1.00	0.77	0.33
$TetX2_{T280A}$ with $K_{M(MCN)}$ = 16.2	1.00	0.81	0.39

Growth rates were calculated using the growth rate function

Effect of Increasing V_{max} by 10% on Growth Rate

	MCN (μM)		
	4	32	64
TetX2 WT with V_{max} = 0.16	0.99	0.24	0.04
TetX2 WT with V_{max} = 0.18	0.99	0.31	0.06
$TetX2_{T280A}$ with V_{max}= 0.15	1.00	0.77	0.33
$TetX2_{T280A}$ with V_{max}= 0.16	1.00	0.81	0.4

Growth rates were calculated using the growth rate function

stage of fitting. Analysis of the shape of the growth rate curves generated from a fixed $V_{max}/K_{M(MCN)}$ ratio with varying absolute vales of V_{max} and $K_{M(MCN)}$ shows that while there is a unique solution for each value of $K_{M(MCN)}$ and V_{max} for a given activity ratio, as the absolute values of $K_{M(MCN)}$ and V_{max} increase (past $K_{M(MCN)} = 100$) the function collapses asymptotically to a maximum curve (figure 5.1B). If the growth rates of a mutant are significantly under predicted when the curve is being fit, there is a tendency for a lack of convergence because the fitting algorithm will continue to increase $K_{M(MCN)}$ without afffecting the shape of the curve. The lack of convergence means the model becomes unstable at this extreme edge. As shown in figure 5.4 and table 5.2 the agreement on the ratio of $V_{max}/K_{M(MCN)}$ was very good when compared with experimentally measured values. Solutions were then sorted by calculating the maximum slope of the experimental growth rate curve (near the midpoint) using 5 points and comparing the slope of the set of predicted growth rate curves at the same MCN concentration (figure 5.1C). We tried a variety of different methods for sorting the solutions but found that using the midpoint functioned best by minimizing the effects of errors in the fit of the initial plateau as well as the tail of the curve by focusing on the linear part of the curve. The ratio ($V_{max}/K_{M(MCN)}$) of the solution from the first set that best fits the slope of the growth rates (figure 5.1) is used as a constraint in a second fit that is used to estimate $K_{M(MCN)}$ (table 5.2). The value for V_{max} was then determined using the predicted activity ratio and $K_{M(MCN)}$. We did not observe a significant difference between the measured and estimated values of $V_{max}/K_{M(MCN)}$ (paired t-test: df = 7, p = 0.2).

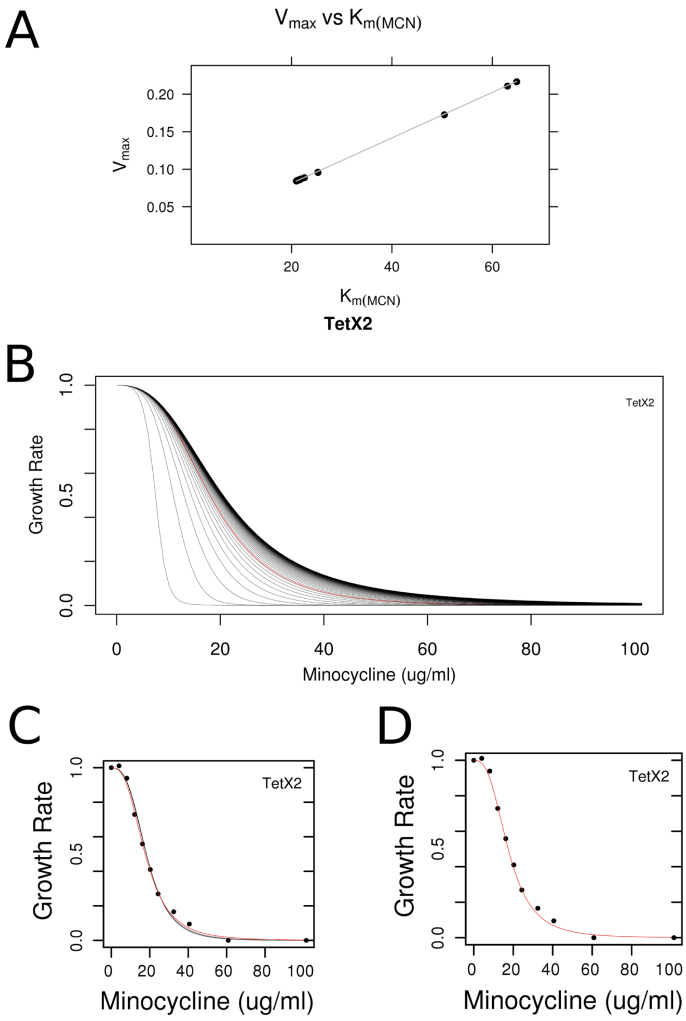


Figure 5.1: A) An initial fit is done to determine the $V_{max}/K_{M(MCN)}$ ratio that provides the best fit for each TetX2 mutant. Fitting with multiple starting conditions results in a series of solutions. The fits with the lowest residuals share the same ratio. B) The same ratio of $V_{max}/K_{M(MCN)}$ results in a series of curves with distinct effects on predicted growth rates. C) The solution that matches the slope of the growth curve best at the midpoint is used to determine the appropriate $V_{max}/K_{M(MCN)}$. D) The growth curve is fit again using the $V_{max}/K_{M(MCN)}$ value determined in the previous step to estimate V_{max} and $K_{M(MCN)}$.

Table 5.2: Comparison of predicted and calculated steady-state kinetic parameters for wild-type TetX2 and adaptive mutants

	measured	predicted	measured	predicted	measured	predicted
	$V_{max}/K_{M(MCN)}$ (μM^{-1} s^{-1})	$V_{max}/K_{M(MCN)}$ (μM^{-1} s^{-1})	$K_{M(MCN)}$ (μM)	$K_{M(MCN)}$ (μM)	V_{max} (s^{-1})	V_{max} (s^{-1})
TetX2	0.005 ± 0.0003	0.004	35 ± 1.9	25.7	0.16 ± 0.005	0.1
T280A	0.011 ± 0.0006	0.019	18 ± 0.9	5.1	0.19 ± 0.004	0.1
N371I	0.013 ± 0.0018	0.013	18 ± 1.9	39.9	0.23 ± 0.015	0.54
N371T	0.008 ± 0.0011	0.011	24 ± 2.1	26.3	0.20 ± 0.013	0.3
S326I	0.007 ± 0.0005	0.007	37 ± 2.8	59.5	0.25 ± 0.007	0.44
F235Y	0.005 ± 0.0009	0.006	54 ± 6.1	64.9	0.28 ± 0.041	0.42
K64R	0.004 ± 0.0006	0.004	36 ± 4.6	13.3	0.16 ± 0.013	0.06
T280S	0.005 ± 0.0008	0.004	30 ± 3.4	13.5	0.16 ± 0.009	0.06

*T280A experimental measurement of $K_{M(MCN)}$ may be underestimated as the assay cannot easily distinguish $K_{M(MCN)}$ <20μM.

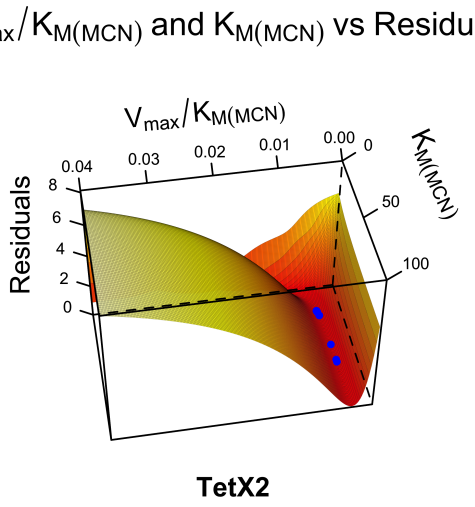
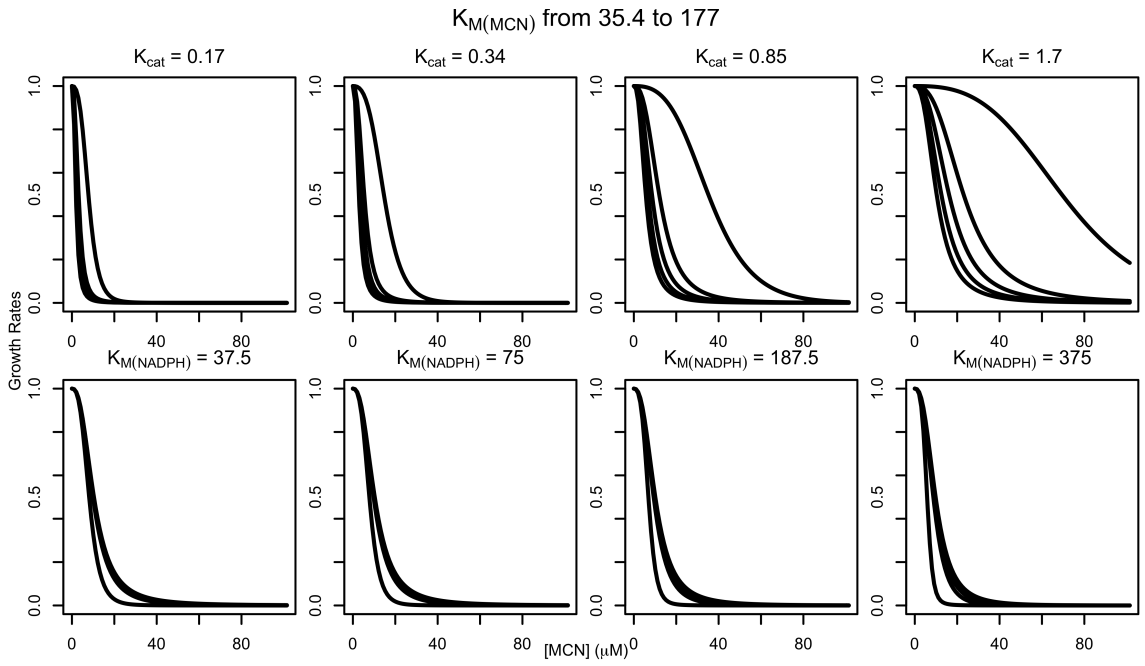


Figure 5.2: Surface of sum squared difference values for the model vs the measured growth rates at different values for parameters $K_{M(MCN)}$ and $V_{max}/K_{M(MCN)}$. The furrow represents the value of $V_{max}/K_{M(MCN)}$ that best fits the shape of the curve. Blue points represent fit solutions for K_M and $V_{max}/K_{M(MCN)}$ using different starting values for the model algorithm.

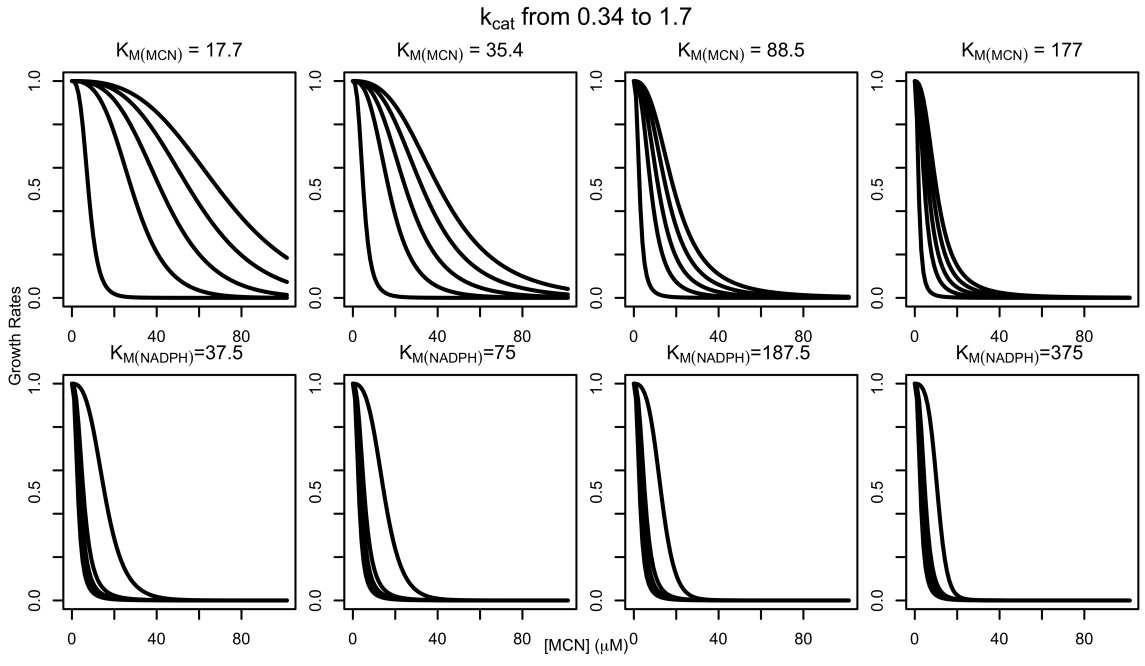
5.3 Jackknife Method for Determining Model Variation

A statistical jackknife method was used to test the variance of the model fit for each of the mutants. Each of the variants of TetX2 was removed from the data set sequentially and fits of the parameters A, B, and D were calculated for the remaining seven variants. The variance of the predicted values was measured by calculating the mean and the standard deviation of the A, B, and D predicted values from each set of parameters (table 5.3). With the exception of F235Y, removing any of the mutants did not have a large effect on the predicted parameters (table 5.3).

We used the parameters predicted from the jackknife method to generate a set of predictions for kinetic parameters for all 8 adaptive mutants. The mathematical

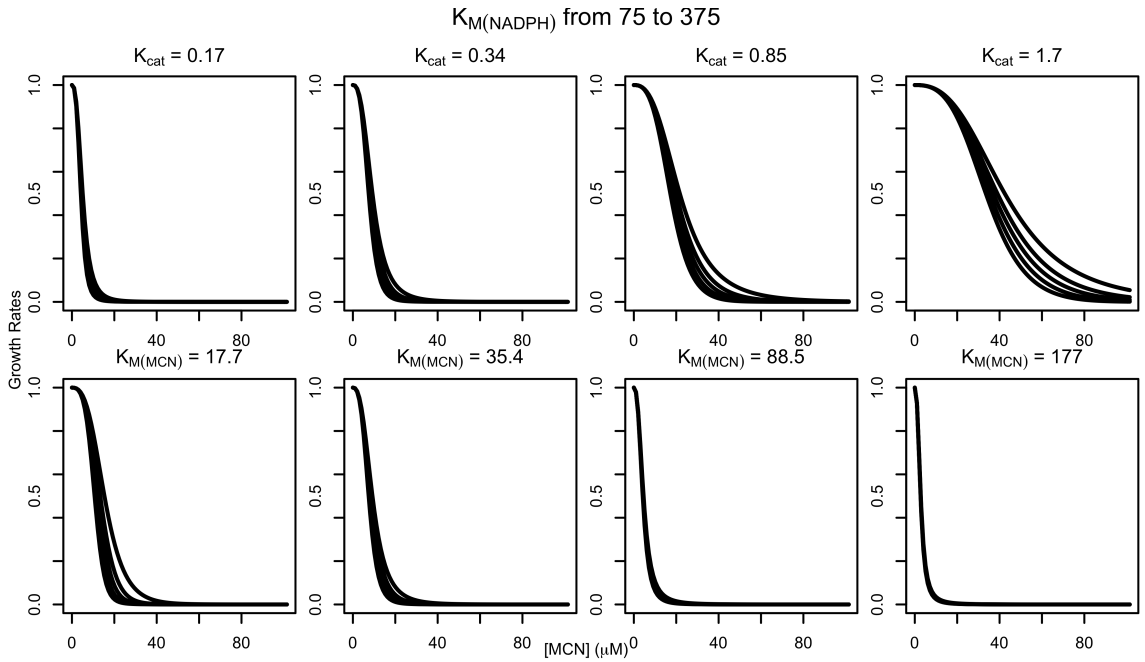

K_M(MCN) from 35.4 to 177
K_cat = 0.17
K_cat = 0.34
K_cat = 0.85
K_cat = 1.7
K_M(NADPH) = 37.5
K_M(NADPH) = 75
K_M(NADPH) = 187.5
K_M(NADPH) = 375
Growth Rates
[MCN] (μM)

(a)


k_cat from 0.34 to 1.7
K_M(MCN) = 17.7
K_M(MCN) = 35.4
K_M(MCN) = 88.5
K_M(MCN) = 177
K_M(NADPH)=37.5
K_M(NADPH)=75
K_M(NADPH)=187.5
K_M(NADPH)=375
Growth Rates
[MCN] (μM)

(b)

Figure 5.3



(c)

Figure 5.3: Changing the kinetic parameters of TetX2 affects the shape of the growth rate curve for MCN resistance. a) Different values of $K_{M(MCN)}$ were used to generate growth rate curves while k_{cat} and $K_{M(NADPH)}$ were held constant at different values. Decreasing $K_{M(MCN)}$ at different values of k_{cat} or $K_{M(NADPH)}$ causes an increase in the initial plateau, increasing the sub inhibitory concentration of MCN for that mutant. b) Changing k_{cat} increases the slope of the growth rate curve but has a lesser effect on the resistance to sub-inhibitory concentrations of MCN. c) A decrease in $K_{M(NADPH)}$ has a similar effect as $K_{M(MCN)}$ but to a much lesser degree. $K_{M(NADPH)}$ therefore has a lesser adaptive role in increasing or decreasing the fitness of the organism.

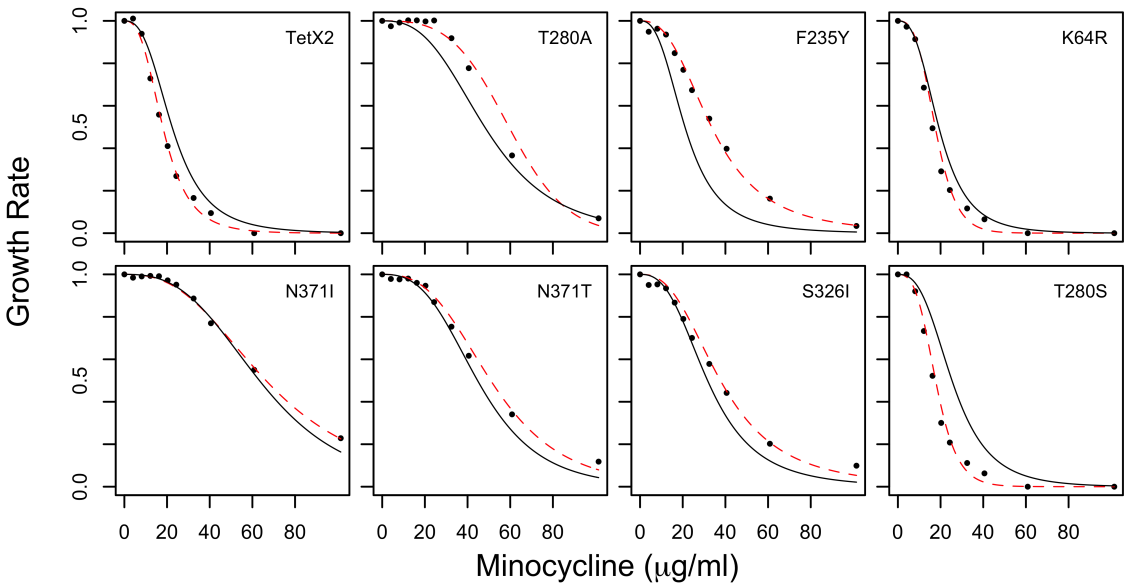


Figure 5.4: A growth rate function built to predict the effect of TetX2 on reducing the inhibition of growth rates by MCN (solid line) was used to predict $K_{M(MCN)}$ and V_{max} for each of the mutants from the measured growth rates (dashed line). We were able to correctly back-calculate the kinetic parameters using our model for each of the mutants (table 5.2).

Table 5.3: Error estimates of Model Fits

	A*	B*	D*	SS**
All Mutants	25.8	2.5	6.7×10^{-4}	1.842
TetX2	57.5	2.3	9.6×10^{-4}	0.173
T280A	40.1	2.2	9.5×10^{-4}	0.346
N371I	51.5	2.4	9.0×10^{-4}	0.024
N371T	71.4	2.3	10.0×10^{-4}	0.052
S326I	21.8	2.6	6.1×10^{-4}	0.095
F235Y	0.1	3.2	1.2×10^{-4}	0.509
K64R	50.3	2.3	8.9×10^{-4}	0.105
T280S	37.2	2.4	7.3×10^{-4}	0.538
Average	41.2	2.5	7.7×10^{-4}	
SD	22.2	0.3	2.9×10^{-4}	

* Error estimates when removing each mutant were done using a Jackknife estimator approach.

** Sum of squared residuals (SS) were measured for each of the individual curves from the mathematical fit using all 8 mutants.

fits for kinetics values of 7 out of the 8 mutants were able to converge to similar values predicted from the parameters of the full fit. The results for predictions of parameters for N371I are shown in table 5.4 as a representative set, with the rest of the predictions in appendix C. However, removing F235Y from the mathematical fit for the parameters causes a failure of the model to successfully converge when predicting the values for F235Y, N371I, and S326I. The model under-predicts growth rates for those mutants, causing the kinetic parameter predictions to fall into the regime where the fit is unable to converge. The mutant F235Y does not show an apparent difference in the shape of the growth rate curve, however, F235Y has a significantly higher $K_{M(MCN)}$ than the other seven mutants as shown in table 4.1. Analysis of the kinetic parameters reveals that we can sort the rest of the adaptive mutants into similar patterns. We theorize that the fit of the growth rate function requires mutants that fall into distinct categories to constrain the model in some form. Adaptive mutants that have growth rate curves or kinetic parameters that fall outside of the range of the original 8 mutants could therefore be very useful in increasing the accuracy of the model.

For five of the seven mutants the growth rate and fitness effects are accounted for largely by kinetic parameters as shown in table 5.2. The experimental data at the lowest values of $K_{M(MCN)}$ may be overestimated because it is difficult to experimentally estimate $K_{M(MCN)}$ when values are less than 20 μM. The model fit for TetX2 growth rates is most sensitive to $K_{M(MCN)}$ and is not as sensitive to $K_{M(NADPH)}$ at the concentrations of MCN where our adaptation experiment occurs. $K_{M(MCN)}$ affects the initial plateau where there is little effect on growth rate when increasing MCN. The measured k_{cat} values were similar for seven of the eight variants and therefore played a lesser role in determining fitness. The total activity affects the slope of the

Table 5.4: Estimate of kinetic parameters for $tetX2_{N371I}$ using jackknife method model parameters

Variant Removed*	$K_{M(MCN)}$ (μM)	V_{max} (s^{-1})	$V_{max}/K_{M(MCN)}$ ($\mu M^{-1}\ s^{-1}$)
TetX2	29.61	0.4	0.014
T280A	32.11	0.46	0.014
F235Y	3.93×10^{6}	5.68×10^{4}	0.014
K64R	41.38	0.53	0.013
N371I	60.93	0.77	0.013
N371T	36.07	0.5	0.014
S326I	53.34	0.72	0.013
T280S	50.17	0.61	0.012

*Variant removed when calculating growth rate function parameters A, B, and D.

curve defining how severely the growth rate decreases in response to increasing MCN. The model cannot discriminate between changes in k_{cat} or changes in active protein concentration since these two variables affect the shape of the growth curve in the same way. Thus the value that is best estimated by the model from growth rates alone is for $V_{max}/K_{M(MCN)}$. Using growth rates alone, the model is able to robustly estimate $V_{max}/K_{M(MCN)}$ which, in turn, can be used to estimate $K_{M(MCN)}$ and V_{max}. The MCN concentration where the most successful adaptive mutant ($TetX2_{T280A}$) was first observed by experimental evolution (32 μM) is very near the range of $K_{M(MCN)}$s for the most successful adaptive TetX2 mutants identified *in vitro* (table 4.1).

5.4 Using a Novel Mutant to Validate the Growth Rate Function

A new double mutant of TetX2, $TetX2_{T280A/N371D}$ was identified from a second random mutagenesis experiment by Anisha Perez. The growth rates and kinetics parameters were measured by Anisha Perez, and we used the growth rate function with the parameter predictions from the first 8 mutants to compare the predictions of $K_{M(MCN)}$ and $V_{max}$to measured values. We were able to accurately predict the $K_{M(MCN)}$ for the mutant (table 5.5). Without a measurement of protein concentration from the whole cell extract we do not know what the V_{max} is for this enzyme. However, we can predict what the protein concentration would have to be to match our predicted V_{max} to the measured k_{cat}. We predict that the protein concentration would have to be between 1.5 and 2.0 times the wild type concentration, which is comparable to some of the higher protein concentration mutants from the initial 8 mutants isolated.

Table 5.5: TetX2$_{T280A/N371S}$ Measured and Predicted Kinetic Parameters

	V_{max} (s^{-1})	k_{cat} (s^{-1})	$K_{M(NADPH)}$ (μM)	$K_{M(MCN)}$ (μM)	$V_{max}/K_{M(MCN)}$ (μM^{-1} s^{-1})
measured	NA	0.94	165.7	47.86	NA
predicted	0.69	NA	NA	67.7	0.01

NA = values not measured

5.5 Conclusions

We were able to reliably predict the growth rates for *E. coli* containing *tetX2*$_{T280A}$, *tetX2*$_{N371I}$, *tetX2*$_{N371T}$, and *tetX2*$_{K64R}$ largely from the kinetic parameters. The growth rates show a strong dependence on $K_{M(MCN)}$ and V_{max} and a lesser effect from $K_{M(NADPH)}$. As expected from the range of selection for MCN and bisubstrate kinetics, the fit for growth rates of *E. coli* with *tetX2* is most sensitive to K_M for MCN and is not as sensitive to K_M for NADPH under these selection conditions. The combination of $K_{M(MCN)}$ and k_{cat} determine the shape of the growth rate curves with the initial plateaus being particularly sensitive to $K_{M(MCN)}$. The total activity affects the slope of the curve defining how severe the drop off in growth rate is to increasing MCN. Since increasing k_{cat} has the same effect on fitness as increasing TetX2 concentration in our model, the parameters are linked in how they affect the shape of the growth rate curve. The model cannot discriminate between changes in k_{cat} or changes in active protein concentration since these variables only appear in combination with one another in the model, and therefore have the same effect on the shape of the growth rate curve. Fitness functions were determined for $K_{M(MCN)}$ and V_{max} using the growth rate function for the MCN concentration range of the adaptation experiment. The fitness function allows us to make predictions about the

effects of mutations that alter kinetic parameters on the fitness of the organism. The growth rate function correctly predicts the correlation between mutations that decrease $K_{M(MCN)}$ and the most successful adaptive mutants, *tetX2*$_{T280A}$ and *tetX2*$_{N371I}$ in the range of the experimental evolution experiment (20 to 32 μM MCN). In population histograms of the MCN adaptation experiments the two mutations *tetX2*$_{T280A}$ and *tetX2*$_{N371I}$ also appear with the highest frequency, illustrating the usefulness of the model on predicting evolutionary outcomes. The growth rate function can also be used to make fast and reliable predictions of enzyme kinetic parameters. Once we have determine the model parameters from a test set of enzyme kinetics and growth rates, the model can be used to reliably predict $V_{max}/K_{M(MCN)}$, and to predict $K_{M(MCN)}$ and $V_{max}/K_{M(MCN)}$ with reasonable accuracy.

Improvements on the growth rate function can increase the accuracy of predictions on kinetic parameters and growth rates. Improvements on the accuracy of measurement of growth rates and enzyme kinetics can improve the sensitivity of the model and possibly allow us to predict the effect of variables that have a smaller effect on fitness, such as $K_{M(NADPH)}$. The growth rate function can also benefit from a better understanding of the effect of the *in vitro* stability of TetX2 on the concentration of TetX2 in the cell at steady state. Two factors that can have a smaller effect on modeling the fitness of TetX2 mutants are the fitness costs of producing TetX2 and the effect of regulation on TetX2 concentration. *TetX2* was inserted in the chromosome in the *spc* operon between *prlA* (SecY) and *rmpJ* (L36), which is tightly regulated, reducing the possibility of regulatory changes on increasing TetX2 concentration. However, mutations in regulatory regions are still possible sources of increased MCN resistance using TetX2 and cannot be easily discounted. The difference in steady state concentrations between TetX2 mutants is sufficient to cause an increase in the fitness cost of

producing the enzyme. However, since selection for mutants with increased resistance is very strong in this experiment, mutations that enhance antibiotic resistance will be strongly favored regardless of associated fitness costs, and the experiment does not last long enough for compensatory mutations to arise. The growth rate function can therefore still be improved by investigating the effect on the growth rate equation of both the fitness cost of protein production and the effect of regulation on protein concentrations. The growth rate function can also be improved by expanding the model to take into account other mechanisms of antibiotic resistance, as well as generalizing the model to function for other bacteriostatic antibiotics.

Chapter 6

Conclusions and Future Work

6.1 Discussion

The role of biophysical properties on determining evolutionary outcomes was investigated using two separate evolution experiments. Reducing the temperature ramp allowed us to observe the rise of triple mutants of adenylate kinase that provide increased activity by improving thermostability. The lack of novel mutation sites in adenylate kinase adaptation under weak selection compared with strong selection suggests that the adaptive landscape for adenylate kinase under our selection conditions is strongly constrained. Relief of selective pressure from the reduced temperature ramp should have lead to the appearance of novel mutations with small or near neutral effects on fitness. Instead we observed that possible adaptive mutants were limited to mutations that had been previously identified under strong selection, indicating that the adaptive landscape is much smaller than we had predicted. Negative epistasis between mutations A193V and T179I with Q16L is consistent with the theory of diminishing returns epistasis, since as the cell approaches a fitness peak,

further increases in stability provide diminishing returns on fitness [Chou et al., 2011]. The stochasticity of the weak selection experiment is also consistent with the expectation that reduced selection strength increases the effect of drift on evolutionary outcomes [Barrett et al., 2006]. However, the temperature ramp of 0.2℃ per day, while significantly weaker than the previously used ramp of 0.5℃ per day, might be in more of a medium selection strength regime as opposed to a truly weak selection strength regime. However, the temperature ramp of 0.2℃ per day, while significantly weaker than the previously used ramp of 0.5℃ per day, might not be weak enough to allow neutral or near neutral mutations to affect the outcome of the evolution experiment significantly. Technical limitations on controlling temperature variation in the vessel prevent us from using a weaker temperature ramp, but if we were able to reduce the temperature ramp the effects of drift can become significantly stronger, and can allow for increased exploration of the adaptive landscape by the bacterial population during adaptation.

The growth rate function relating enzyme properties to growth rates is able to provide a direct link between physicochemical changes and fitness. We were able to create a model that can predict fitness benefits from physicochemical properties of enzymes and conversely, can predict enzyme properties from growth rates measurements. The model is also able to predict numerical values for fitness changes from changes to enzyme properties. Surprisingly the benefit from decreasing the Michaelis constant K_M for NADPH does not have a strong effect on fitness outcomes compared with K_M for MCN under normal cell conditions. Strong regulation of NADPH concentration by cellular metabolism, and the already low K_M for NADPH of wild type TetX2 compared with the cellular concentration, might suggest why there is little benefit to be gained from decreasing $K_{M(NADPH)}$. Although changes in $K_{M(NADPH)}$

were observed, there is no clear correlation between successful mutants and decreased $K_{M(NADPH)}$, suggesting that the fitness prediction of the model is correct. If there was a decrease in $K_{M(NADPH)}$ in response to MCN inhibition, such that enzyme efficiency was affected, the fitness benefit of decreasing $K_{M(NADPH)}$ would be expected to increase, leading to adaptive mutants with increased affinity for NADPH. One of the limitations of the present model is the failure to consider the cost of enzyme production on fitness [Lenski, 1998], however, since the model is still accurate without this added constraint the fitness cost of producing antibiotic resistance enzymes probably does not play a strong role in determining fitness under the very strong antibiotic inhibition of growth rates of our experiment. If selection strength is decreased, there is a probability that fitness costs of producing these enzymes can become more relevant. However, under our model, both increasing enzyme activity and increasing enzyme concentration provide equivalent fitness benefits, and their effect on enzyme rate is multiplicative.

The accurate prediction of the $V_{max}/K_{M(MCN)}$ ratio using the growth rate function provides a very useful and high-throughput method for determining enzyme properties from growth rates studies. Calculation of kinetic parameters from flux model analysis removes the requirement for protein purification, which can be expensive in both time and effort. A set of source growth rates and enzyme kinetic measurements with distinct properties is still required to provide an initial model. Once the initial fit is populated, enzyme properties can be predicted, using only information of growth rates, to quickly identify mutants with interesting enzyme properties. Candidate enzymes that are predicted to have improved enzyme efficiency can then be isolated and purified for more in-depth measurements of enzyme kinetics. The model we have developed can be used with other enzymes that degrade bacteriostatic antibiotics. The

model can also be used with enzymes that provide antibiotic resistance through other mechanisms, or combinations of enzymes, by expanding the mathematical equation used to predict the steady state MCN_C concentration.

6.2 Future Work: Extending the Growth Rate Function to Enzyme Pathways

We are interested in extending our model to mulit-enzyme systems by developing a more general equation for the mathematical modeling of growth rates. Of the two other major antibiotic resistance mechanisms, efflux pumps are a compelling target for extending the model. Antibiotic resistance through efflux pumps is the most common form of antibiotic resistance in multi drug resistant bacteria, so creating a model that can allow us to easily compare enzyme parameters without extensive purification would be useful in clinical applications. Efflux pumps work in parallel to enzymatic inactivation to reduce the concentration of cytosolic antibiotic. A suitable target for implementing in the mathematical model is the efflux pump adeABC from *Acinetobacter baumannii*, which confers resistance to the tetracycline derivative tigecycline [Peleg et al., 2007]. The mathematical model described in Thanassi et al., [Thanassi et al., 1995] also provides a useful mathematical equation for the kinetics of an efflux pump. We can use Michaelis-Menten to model the kinetics of the pump in the same manner as for TetX2. Diffusion and accumulation of antibiotic in the cell is unchanged since the efflux pump adeABC pumps antibiotic directly into the media, bypassing the periplasm. The steady state concentration of antibiotic in the cytoplasm can be calculated by setting the diffusion rate equal to the sum of the effects of the pump and the inactivating enzyme. Growth rates can be calculated

by using the Hill equation (2.25). The parameters A, B, and D describe interactions between the antibiotic and the ribosome and so are independent of the action of the pump and the inactivating enzyme. Values for A, B, and D determined using the current model are directly applicable if the same bacterial system is extended by the addition of an efflux pump.

6.3 Conclusions

The biophysical and physicochemical properties of enzymes determining fitness outcomes were explored using a combination of experimental evolution and biophysical studies. The integration of top down adaptation experiments with bottom up biophysical studies is necessary to allow a complete understanding of mechanisms that drive fitness differences and adaptation. Evolution experiments can provide a source for enzyme variety needed to investigate the link between physical changes and fitness, but more importantly, allow us to control the environment under which evolution occurs. Control of selection conditions allows us to correlate evolutionary outcomes to specific environmental changes. Accurate predictions of evolutionary outcomes of adaptation experiments from biophysical measurements relies on both correlating physicochemical changes to fitness outcomes, and knowledge of the shape of the adaptive landscape under which fitness differences are acted on. The development of growth rate functions based on flux analysis provides a good example of how metabolic flux can be used to predict fitness benefits. The canonical use of flux balance analysis is to predict growth rates from flux of a limiting resource, but we can also use the same type of analysis to derive a growth rate function based on antibiotic inhibition of growth rates. Evolution experiments also reveal the com-

plexity of evolutionary outcomes, which rely on stochastic processes and epistatic interactions that are not necessarily predicted from measurements of enzyme kinetic properties. Improvements on measurements of kinetic properties and growth rates, coupled with high throughput sequencing methods for determine population distributions of adaptation experiments, will allow us to refine models of the mechanisms that drive adaptation. The relationship between physical properties and fitness can be used to predict the effect of novel mutations that confer antibiotic resistance, as well as understanding the actual physiochemical properties that provide resistance. Coupled with structural and functional studies of mutant enzymes we can approach a more complete understanding of the pathway from structural changes in enzymes to phenotypic changes of organisms.

Bibliography

Alekshun, M. N. and Levy, S. B. (2007). Molecular mechanisms of antibacterial multidrug resistance. Cell *128*, 1037–1050.

Andersen, K. B. and von Meyenburg, K. (1977). Charges of nicotinamide adenine nucleotides and adenylate energy charge as regulatory parameters of the metabolism in Escherichia coli. The Journal of biological chemistry *252*, 4151–4156.

Argast, M. and Beck, C. (1984). Tetracycline diffusion through phospholipid bilayers and binding to phospholipids. Antimicrobial agents and chemotherapy *26*, 263.

Argast, M. and Beck, C. (1985). Tetracycline uptake by susceptible Escherichia coli cells. Archives of microbiology *141*, 260–265.

Arjan G, J., Zeyl, C. W., Gerrish, P. J., Blanchard, J. L. and Lenski, R. E. (1999). Diminishing returns from mutation supply rate in asexual populations. Science *283*, 404–406.

Arnold, F. H., Wintrode, P. L., Miyazaki, K. and Gershenson, A. (2001). How enzymes adapt: lessons from directed evolution. Trends in Biochemical Sciences *26*, 105–106.

Atkinson, D. E. (1968). Energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers. Biochemistry *7*, 4030–4034.

Atkinson, D. E. and Walton, G. M. (1967). Adenosine triphosphate conservation in metabolic regulation. Rat liver citrate cleavage enzyme. The Journal of biological chemistry *242*, 3239–3241.

Bae, E. and Phillips Jr, G. N. (2004). Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. Journal of Biological Chemistry *279*, 28202–28208.

Barlow, M. and Hall, B. G. (2002). Predicting Evolutionary Potential In Vitro Evolution Accurately Reproduces Natural .... Genetics *160*, 823–832.

Barrett, R. D. H., M'gonigle, L. K. and Otto, S. P. (2006). The Distribution of Beneficial Mutant Effects Under Strong Selection. Genetics *174*, 2071–2079.

Bentley, D., Balasubramanian, S., Swerdlow, H., Smith, G., Milton, J., Brown, C., Hall, K., Evers, D., Barnes, C. and Bignell, H. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature *456*, 53–59.

Bjorkholm, B., Sjolund, M., Falk, P. G., Berg, O. G., Engstrand, L. and Andersson, D. I. (2001). Mutation frequency and biological cost of antibiotic resistance in Helicobacter pylori. Proceedings of the National Academy of Sciences *98*, 14607–14612.

Bloom, J. D., Labthavikul, S. T., Otey, C. R. and Arnold, F. H. (2006). Protein stability promotes evolvability. Proceedings of the National Academy of Sciences *103*, 5869–5874.

Bloom, J. D., Silberg, J. J., Wilke, C. O., Drummond, D. A., Adami, C. and Arnold, F. H. (2005). Thermodynamic prediction of protein neutrality. Proceedings of the National Academy of Sciences *102*, 606–611.

Blount, Z. D., Borland, C. Z. and Lenski, R. E. (2008). ... contingency and the evolution of a key innovation in an experimental population of Escherichia coli. Proceedings of the National Academy of Sciences *105*, 7899–7906.

Boucher, H. W., Talbot, G. H., Bradley, J. S., Edwards, J. E., Gilbert, D., Rice, L. B., Scheld, M., Spellberg, B. and Bartlett, J. (2009). Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America *48*, 1–12.

Brodersen, D. E., Clemons, W. M., Carter, A. P., Morgan-Warren, R. J., Wimberly, B. T. and Ramakrishnan, V. (2000). The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. Cell *103*, 1143–1154.

Campos, P. R. A. and Wahl, L. M. (2009). THE EFFECTS OF POPULATION BOTTLENECKS ON CLONAL INTERFERENCE, AND THE ADAPTATION EFFECTIVE POPULATION SIZE. Evolution *63*, 950–958.

Camps, M., Naukkarinen, J., Johnson, B. and Loeb, L. (2003). Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. Proceedings of the National Academy of Sciences *100*, 9727–9732.

Charusanti, P., Fong, N. L., Nagarajan, H., Pereira, A. R., Li, H. J., Abate, E. A., Su, Y., Gerwick, W. H. and Palsson, B. O. (2012). Exploiting adaptive laboratory

evolution of Streptomyces clavuligerus for antibiotic discovery and overproduction. PloS one *7*, e33727.

Chattopadhyay, S., Weissman, S., Minin, V., Russo, T., Dykhuizen, D. and Sokurenko, E. (2009). High frequency of hotspot mutations in core genes of Escherichia coli due to short-term positive selection. Proceedings of the National Academy of Sciences *106*, 12412.

Chopra, I. and Roberts, M. (2001). Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiology and Molecular Biology Reviews *65*, 232.

Chopra, I. and Shales, S. (1981). Susceptibility of protein synthesis in Escherichia coli to tetracycline and minocycline. Journal of general microbiology *124*, 187–189.

Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segre, D. and Marx, C. J. (2011). Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. Science *332*, 1190–1192.

Chubiz, M. L., Lee, M. C., Delaney, N. and Marx, C. (2012). FREQ-Seq: A rapid, cost-effective, sequencing-based method to determine allele frequencies directly from mixed populations. PLoS ONE.

Cleland, W. W. (1963). The kinetics of enzyme-catalyzed reactions with two or more substrates or products. I. Nomenclature and rate equations., vol. 67,. Biochimica et Biophysica Acta.

Cohen, M. L. (2000). Changing patterns of infectious disease. Nature *406*, 762–767.

Collins, S., de Meaux, J. and Acquisti, C. (2007). Adaptive walks toward a moving optimum. Genetics *176*, 1089–1099.

Connell, S. R., Tracz, D. M., Nierhaus, K. H. and Taylor, D. E. (2003). Ribosomal protection proteins and their mechanism of tetracycline resistance. Antimicrobial agents and chemotherapy *47*, 3675–3681.

Cooper, T. F. and Lenski, R. E. (2010). Experimental evolution with E. coli in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. BMC evolutionary biology *10*, 11.

Couñago, R., Chen, S. and Shamoo, Y. (2006). In vivo molecular evolution reveals biophysical origins of organismal fitness. Molecular Cell *22*, 441–449.

Couñago, R. and Shamoo, Y. (2005). Gene replacement of adenylate kinase in the gram-positive thermophile Geobacillus stearothermophilus disrupts adenine nucleotide homeostasis and reduces cell viability. Extremophiles *9*, 135–144.

Couñago, R., Wilson, C. J., Peña, M. I., Wittung-Stafshede, P. and Shamoo, Y. (2008). An adaptive mutation in adenylate kinase that increases organismal fitness is linked to stability-activity trade-offs. Protein engineering, design & selection : PEDS *21*, 19–27.

D E Dykhuizen, D. L. H. (1983). Selection in chemostats. Microbiological reviews *47*, 150.

Darwin, C. R. (1859). On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London: John Murray.

Dean, A. M. and Thornton, J. W. (2007). Mechanistic approaches to the study of evolution: the functional synthesis. Nature Reviews Genetics *8*, 675–688.

Dekel, E. and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. Nature *436*, 588–592.

Denamur, E. and Matic, I. (2006). Evolution of mutation rates in bacteria. Molecular Microbiology *60*, 820–827.

DePristo, M. A., Weinreich, D. M. and Hartl, D. L. (2005). Missense meanderings in sequence space: a biophysical view of protein evolution. Nature Reviews Genetics *6*, 678–687.

Draghi, J. A., Parsons, T. L., Wagner, G. P. and Plotkin, J. B. (2010). Mutational robustness can facilitate adaptation. Nature *463*, 353–355.

Dykhuizen, D. E. and Dean, A. M. (1990). Enzyme activity and fitness: evolution in solution. Trends in Ecology & Evolution *5*, 251–262.

Dykhuizen, D. E., Dean, A. M. and Hartl, D. L. (1987). Metabolic flux and fitness. Genetics *115*, 25–31.

Eanes, W. F. (1987). Allozymes and fitness: Evolution of a problem. Trends in Ecology & Evolution *2*, 44–48.

Easterby, J. (1981). A generalized theory of the transition time for sequential enzyme reactions. Biochemical Journal *1981*, 155–161.

Ecker, R. E. and Schaechter, M. (1963). RIBOSOME CONTENT AND THE RATE OF GROWTH OF SALMONELLA TYPHIMURIUM. Biochimica et biophysica acta *76*, 275–279.

Elena, S. F. and Lenski, R. E. (2003). Microbial genetics: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nature Reviews Genetics *4*, 457–469.

Entsch, B. and van Berkel, W. J. (1995). Structure and mechanism of para-hydroxybenzoate hydroxylase. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *9*, 476–483.

Erlich, Y., Chang, K., Gordon, A., Ronen, R., Navon, O., Rooks, M. and Hannon, G. J. (2009). DNA Sudoku–harnessing high-throughput sequencing for multiplexed specimen analysis. Genome Research *19*, 1–12.

Eyre-Walker, A. and Keightley, P. D. (2007). The distribution of fitness effects of new mutations. Nature Reviews Genetics *8*, 610–618.

File, T. M. (2006). Clinical implications and treatment of multiresistant Streptococcus pneumoniae pneumonia. Clinical Microbiology & Infection *12*, 31–41.

Fisher, R. A. (1930). The genetical theory of natural selection. Oxford: The Clarendon press.

Galhardo, R. S., Hastings, P. J. and Rosenberg, S. M. (2007). Mutation as a stress response and the regulation of evolvability. Critical reviews in biochemistry and molecular biology *42*, 399–435.

Gillespie, J. H. (1986). Natural selection and the molecular clock. Molecular Biology and Evolution *3*, 138–155.

Giske, C. G., Monnet, D. L., Cars, O., Carmeli, Y. and ReAct-Action on Antibiotic Resistance (2008). Clinical and economic impact of common multidrug-resistant gram-negative bacilli. Antimicrobial agents and chemotherapy *52*, 813–821.

Glaser, P., Presecan, E. and Delepierre, M. (1992). Zinc, a novel structural element found in the family of bacterial adenylate kinases. Biochemistry *31*, 3038–3043.

Gonzalez, C., Hadany, L., Ponder, R. G., Price, M., Hastings, P. J. and Rosenberg, S. M. (2008). Mutability and importance of a hypermutable cell subpopulation that produces stress-induced mutants in Escherichia coli. PLoS Genetics *4*, e1000208.

Gould, S. J. and Eldredge, N. (1977). Punctuated equilibria: the tempo and mode of evolution reconsidered. Paleobiology *3*, 115–151.

Hardie, D. G. and Hawley, S. A. (2001). AMP-activated protein kinase: the energy charge hypothesis revisited. BioEssays *23*, 1112–1119.

Hartl, D. L., Dykhuizen, D. E. and Dean, A. M. (1985). Limits of adaptation: the evolution of selective neutrality. Genetics *111*, 655–675.

Ibarra, R. U., S, E. J. and Palsson, B. O. (2002). Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature *420*, 186–189.

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. Journal of computational and graphical statistics *5*, 299–314.

Kawecki, T. J., Lenski, R. E., Ebert, D., Hollis, B., Olivieri, I. and Whitlock, M. C. (2012). The value of complementary approaches in evolutionary research: reply to Magalhães and Matos. Trends in Ecology & Evolution *27*, 547–560.

Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E. and Cooper, T. F. (2011). Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. Science *332*, 1193–1196.

Kimura, M. (1968). Evolutionary rate at the molecular level. Nature *217*, 624–626.

Kornberg, A. and Pricer Jr, W. E. (1951). Enzymatic phosphorylation of adenosine and 2, 6-diaminopurine riboside. The Journal of Biological Chemistry *193*, 481–495.

Kryazhimskiy, S., Tkačik, G. and Plotkin, J. B. (2009). The dynamics of adaptation on correlated fitness landscapes. Proceedings of the National Academy of Sciences *106*, 18638–18643.

Lee, H., Popodi, E., Tang, H. and Foster, P. L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. Proceedings of the National Academy of Sciences *109*, E2774–83.

Lenski, R. E. (1998). Bacterial evolution and the cost of antibiotic resistance. International Microbiology *1*, 265–270.

Lenski, R. E., Rose, M. R., Simpson, S. C. and Tadler, S. C. (1991). Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 Generations. American Naturalist *138*, 1315–1341.

Lerman, J. A., Hyduke, D. R., Latif, H., Portnoy, V. A., Lewis, N. E., Orth, J. D., Schrimpe-Rutledge, A. C., Smith, R. D., Adkins, J. N., Zengler, K. and Palsson, B. O. (2012). In silico method for modelling metabolism and gene product expression at genome scale. Nature communications *3*, 929.

Levin, B. R., Perrot, V. and Walker, N. (2000). Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. Genetics *154*, 985–997.

Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D., Weitz, K. K.,

Eils, R., K Ouml Nig, R., Smith, R. D. and Palsson, B. O. (2010). Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Molecular Systems Biology *6*, 1–13.

Livermore, D. M. (2003). Bacterial resistance: origins, epidemiology, and impact. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America *36*, S11–23.

Lunzer, M., Golding, G. B. and Dean, A. M. (2010). Pervasive cryptic epistasis in molecular evolution. PLoS Genetics *6*, e1001162.

Lunzer, M., Natarajan, A., Dykhuizen, D. E. and Dean, A. M. (2002). Enzyme kinetics, substitutable resources and competition: from biochemistry to frequency-dependent selection in lac. Genetics *162*, 485–499.

McMurry, L. M., Cullinane, J. and Levy, S. (1982). Transport of the lipophilic analog minocycline differs from that of tetracycline in susceptible and resistant Escherichia coli strains. Antimicrobial agents and chemotherapy *22*, 791.

Miller, S. P., Lunzer, M. and Dean, A. M. (2006). Direct demonstration of an adaptive constraint. Science *314*, 458–461.

Moore, I. F., Hughes, D. W. and Wright, G. D. (2005). Tigecycline is modified by the flavin-dependent monooxygenase TetX. Biochemistry *44*, 11829–11835.

Müller, C. W., Schlauderer, G. J., Reinstein, J. and Schulz, G. E. (1996). Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. Structure *4*, 147–156.

Nagel, A. C., Joyce, P., Wichman, H. A. and Miller, C. R. (2012). Stickbreaking: a novel fitness landscape model that harbors epistasis and is consistent with commonly observed patterns of adaptive evolution. Genetics *190*, 655–667.

Newcomb, R. D., Campbell, P. M., Ollis, D. L., Cheah, E., Russell, R. J. and Oakeshott, J. G. (1997). A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. Proceedings of the National Academy of Sciences of the United States of America *94*, 7464–7468.

Nikaido, H. and Thanassi, D. G. (1993). Penetration of lipophilic agents with multiple protonation sites into bacterial cells: tetracyclines and fluoroquinolones as examples. Antimicrobial agents and chemotherapy *37*, 1393–1399.

Oelschlaeger, P. (2008). Outsmarting metallo-?-lactamases by mimicking their natural evolution. Journal of Inorganic Biochemistry *102*, 2043–2051.

Orencia, M. C., Yoon, J. S., Ness, J. E., Stemmer, W. P. C. and Stevens, R. C. (2001). Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. Nature Structural & Molecular Biology *8*, 238–242.

Orr, H. (2009). Fundamental concepts in genetics: Fitness and its role in evolutionary genetics. Nature Reviews Genetics *10*, 531–539.

Orr, H. A. (2005). The genetic theory of adaptation: a brief history. Nature Reviews Genetics *6*, 119–127.

Orr, H. A. (2006). The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation. Journal of theoretical biology *238*, 279–285.

Orth, J. D., Thiele, I. and Palsson, B. O. (2010). What is flux balance analysis? Nature Reviews Microbiology *28*, 245–248.

Parolo, M., Avena, M., Pettinari, G., Zajonkovsky, I., Valles, J. and Baschini, M. (2010). Antimicrobial properties of tetracycline and minocycline-montmorillonites. Applied Clay Science *49*, 194–199.

Peleg, A. Y., Adams, J. and Paterson, D. L. (2007). Tigecycline Efflux as a Mechanism for Nonsusceptibility in Acinetobacter baumannii. Antimicrobial agents and chemotherapy *51*, 2065–2069.

Peña, M. I., Davlieva, M., Bennett, M. R., Olson, J. S. and Shamoo, Y. (2011). Evolutionary fates within a microbial population highlight an essential role for protein folding during natural selection. Molecular Systems Biology *6*, 1–11.

Peña, M. I., Van Itallie, E., Bennett, M. R. and Shamoo, Y. (2010). Evolution of a single gene highlights the complexity underlying molecular descriptions of fitness. Chaos (Woodbury, N.Y.) *20*, 026107.

Perfeito, L., Ghozzi, S., Berg, J., Schnetz, K. and Lässig, M. (2011). Nonlinear fitness landscape of a molecular pathway. PLoS Genetics *7*, e1002160.

Pioletti, M., Schlünzen, F., Harms, J., Zarivach, R., Glühmann, M., Avila, H., Bashan, A., Bartels, H., Auerbach, T., Jacobi, C., Hartsch, T., Yonath, A. and Franceschi, F. (2001). Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. The EMBO journal *20*, 1829–1839.

Poyart, C., Mugnier, P., Quesne, G., Berche, P. and Trieu-Cuot, P. (1998). A novel extended-spectrum TEM-type beta-lactamase (TEM-52) associated with decreased

susceptibility to moxalactam in Klebsiella pneumoniae. Antimicrobial agents and chemotherapy *42*, 108–113.

Ragoussis, J. (2006). Genotyping technologies for all. Drug Discovery Today: Technologies *3*, 115–122.

Rendel, M. D. (2010). Adaptive evolutionary walks require neutral intermediates in RNA fitness landscapes. Theoretical population biology *79*, 12–18.

Roovers, M., Sanchez, R., Legrain, C. and Glansdorff, N. (2001). Experimental evolution of enzyme temperature activity profile: selection in vivo and characterization of low-temperature-adapted mutants of Pyrococcus furiosus ornithine carbamoyltransferase. Journal of Bacteriology *183*, 1101–1105.

Rossell, S., Solem, C., Jensen, P. R. and Heijnen, J. J. (2011). Towards a quantitative prediction of the fluxome from the proteome. Metabolic engineering *13*, 253–262.

Saint Girons, I., Gilles, A. and Margarita, D. (1987). Structural and catalytic characteristics of Escherichia coli adenylate kinase. Journal of Biological Chemistry *262*, 622–629.

Sasaki, A. and Nowak, M. (2003). Mutation landscapes. Journal of theoretical biology *224*, 241–247.

Serrano, L., Day, A. G. and Fersht, A. R. (1993). Step-wise Mutation of Barnase to Binase. Journal of molecular biology *233*, 305–312.

Sigler, A., Schubert, P., Hillen, W. and Niederweis, M. (2000). Permeation of tetracyclines through membranes of liposomes and Escherichia coli. European journal of biochemistry / FEBS *267*, 527–534.

Smith, J. M. (1970). Natural selection and the concept of a protein space. Nature *225*, 563–564.

Soskine, M. and Tawfik, D. S. (2010). Mutational effects and the evolution of new protein functions. Nature Reviews Genetics *11*, 572–582.

Spahn, C. M. and Prescott, C. D. (1996). Throwing a spanner in the works: antibiotics and the translation apparatus. Journal of molecular medicine (Berlin, Germany) *74*, 423–439.

Speer, B. S., Bedzyk, L. and Salyers, A. A. (1991). Evidence that a novel tetracycline resistance gene found on two Bacteroides transposons encodes an NADP-requiring oxidoreductase. Journal of Bacteriology *173*, 176–183.

Spellberg, B., Guidos, R., Gilbert, D., Bradley, J., Boucher, H. W., Scheld, W. M., Bartlett, J. G., Edwards, J. and the Infectious Diseases Society of America (2008). The Epidemic of Antibiotic-Resistant Infections: A Call to Action for the Medical Community from the Infectious Diseases Society of America. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America *46*, 155–164.

Stemmer, W. P. C. (1994). Rapid evolution of a protein in vitro by DNA shuffling. Nature *370*, 389–390.

Stepanova, M. N., Pimkin, M., Nikulin, A. A., Kozyreva, V. K., Agapova, E. D. and Edelstein, M. V. (2008). Convergent In Vivo and In Vitro Selection of Ceftazidime Resistance Mutations at Position 167 of CTX-M-3 -Lactamase in Hypermutable Escherichia coli Strains. Antimicrobial agents and chemotherapy *52*, 1297–1301.

Storz, J. F. and Wheat, C. W. (2010). Integrating evolutionary and functional approaches to infer adaptation at specific loci. Evolution *64*, 2489–2509.

Swedes, J. S., Sedo, R. J. and Atkinson, D. E. (1975). Relation of growth and protein synthesis to the adenylate energy charge in an adenine-requiring mutant of Escherichia coli. The Journal of biological chemistry *250*, 6930–6938.

Tanaka, M. M., Bergstrom, C. T. and Levin, B. R. (2003). The evolution of mutator genes in bacterial populations: the roles of environmental change and timing. Genetics *164*, 843–854.

Taverna, D. M. and Goldstein, R. A. (2002). Why are proteins so robust to site mutations? Journal of molecular biology *315*, 479–484.

Thanassi, D. G., Suh, G. S. and Nikaido, H. (1995). Role of outer membrane barrier in efflux-mediated tetracycline resistance of Escherichia coli. Journal of Bacteriology *177*, 998–1007.

Tokuriki, N., Stricher, F., Serrano, L. and Tawfik, D. S. (2008). How protein stability and new functions trade off. PLoS computational biology *4*, 1–7.

Travisano, M., Mongold, J. A., Bennett, A. F. and Lenski, R. E. (1995). Experimental tests of the roles of adaptation, chance, and history in evolution. Science *267*, 87–90.

Vakulenko, S., Geryk, B., Kotra, L., Shahriar, M. and Lerner, S. A. (1998). Selection and characterization of ?-lactam-?-lactamase inactivator-resistant mutants .... Antimicrobial agents and chemotherapy *42*, 1542–1548.

Vonrhein, C., Schlauderer, G. J. and Schulz, G. E. (1995). Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. Structure *3*, 483–490.

Walkiewicz, K., tez Cárdenas, A. S. B., Sun, C., Bacorn, C., Saxer, G. and Shamoo, Y. (2012). Small changes in enzyme function can lead to surprisingly large fitness effects during adaptive evolution of antibiotic resistance. PNAS.

Wang, X., Minasov, G. and Shoichet, B. K. (2002). Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. Journal of molecular biology *320*, 85–95.

Weinreich, D. M., Delaney, N. F., DePristo, M. A. and Hartl, D. L. (2006). Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. Science *312*, 111–114.

Weiss, J. N. (1997). The Hill equation revisited: uses and misuses. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *11*, 835–841.

Whittle, G., Hund, B. D., Shoemaker, N. B. and Salyers, A. A. (2001). Characterization of the 13-kilobase ermF region of the Bacteroides conjugative transposon CTnDOT. Applied and environmental microbiology *67*, 3488–3495.

Wichman, H., Scott, L., Yarber, C. and Bull, J. (2000). Experimental evolution recapitulates natural evolution. Philosophical Transactions of the Royal Society B: Biological Sciences *355*, 1677–1684.

Wielgoss, S., Barrick, J. E., Tenaillon, O., Cruveiller, S., Chane-Woon-Ming, B., Médigue, C., Lenski, R. E. and Schneider, D. (2011). Mutation Rate Inferred From

Synonymous Substitutions in a Long-Term Evolution Experiment With Escherichia coli. G3: Genes *1*, 183–186.

Wood, T., Burke, J. and Rieseberg, L. (2005). Parallel genotypic adaptation: when evolution repeats itself. Genetica *123*, 157–170.

Wright, S. (1988). Surfaces of selective value revisited. American Naturalist *131*, 115–123.

Yang, W., Moore, I. F., Koteva, K. P., Bareich, D. C., Hughes, D. W. and Wright, G. D. (2004). TetX is a flavin-dependent monooxygenase conferring resistance to tetracycline antibiotics. The Journal of biological chemistry *279*, 52346–52352.

Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E. and Kim, J. F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature *461*, 1243–1247.

Zarr, J. H. (1999). Biostatistical analysis. New Jersey: Prentice Hall.

Zhang, Q., Robin, K., Liao, D., Lambert, G. and Austin, R. H. (2011). The Goldilocks Principle and Antibiotic Resistance in Bacteria. Molecular Pharmaceutics *8*, 2063–2068.

Appendix A

Model Output Figures

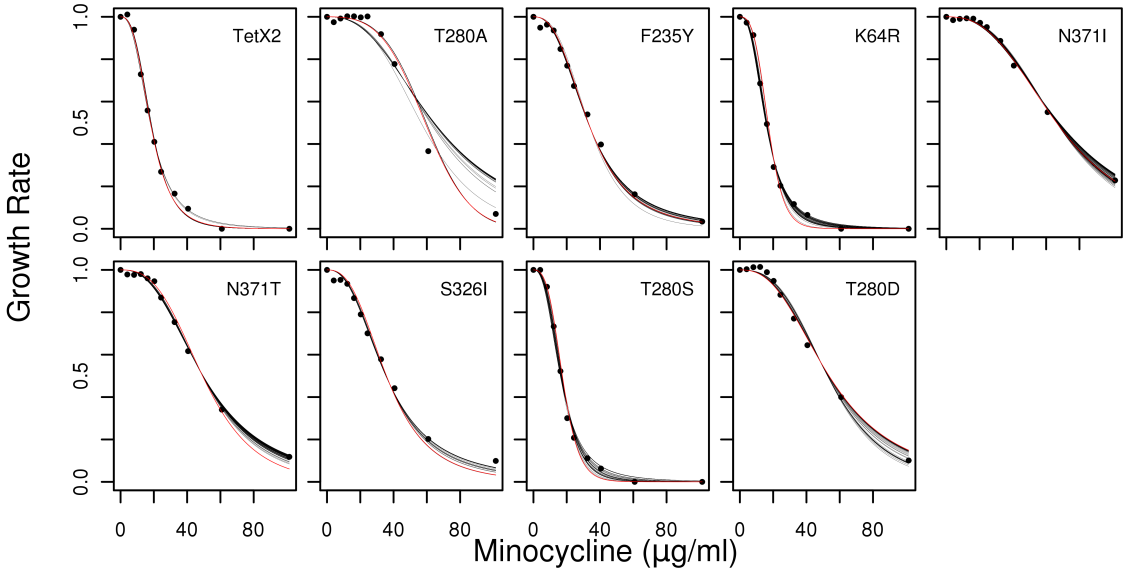


Figure A.1: Solutions from the first step of the mathematical modeling for all 9 mutants, including the T280A/N371S double mutant (labeled T280D)

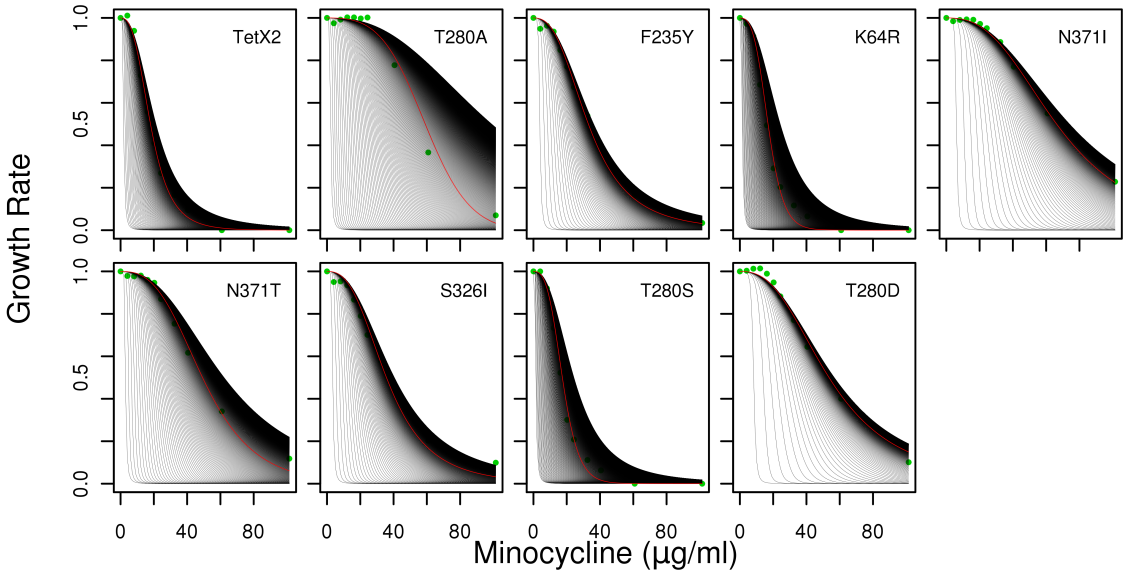


Figure A.2: The effect of varying $K_{M(MCN)}$ with a constant $V_{max}/K_{M(MCN)}$ determined by the first step of the enzyme parameter fit on growth rate curves.

Appendix B

MATLAB and R Scripts

B.1 MATLAB script for modeling coupled assay

The MATLAB script that was used to model the coupled assay is presented here verbatim.

```
function [ATP,AMP,ADP,PEP,PYR,NADH] = CoupAssayBi_PKbisubstrate(ATP,AMP,
    ADP,PEP,PYR,NADH,AK_Vmax,ATP_Km,AMP_Km,PK_Vmax,PK_Km,LDH_Vmax,PYR_Km
    ,NADH_Km,n,l)

  time = 0;

% Michaelis Menten Equations for AK, PK, LDH reactions assuming single
    substrate kinetics
% for AK v = vmax*[ATP]*[AMP]/(Km_ATP*Km_AMP+Km_AMP*[ATP]+Km_ATP*[AMP])
    function [dATP] = AK(ATP,AMP,AK_Vmax,ATP_Km,AMP_Km)
    dATP = (AK_Vmax*ATP*AMP)/(ATP_Km*AMP_Km+AMP_Km*ATP+ATP_Km*AMP);
    end

% for PK v = vmax*[ADP]*[PEP]/(Km_PEP*Km_ADP+Km_PEP*[ADP]+Km_PEP*[ADP])
```

```
    function [dADP] = PK(ADP,PEP,PK_Vmax,PK_Km)
13  dADP = (PK_Vmax*ADP*PEP)/(PK_Km*PK_Km+PK_Km*PEP+PK_Km*ADP);
    end
15

% for LDH v = vmax*[Pyruvate]*[NADH]/(Km_Pyruvate*Km_NADH+Km_Pyruvate*[
    NADH]+Km_NADH*[Pyruvate])
17  function [dPYR] = LDH(PYR,NADH,LDH_Vmax,PYR_Km,NADH_Km)
    dPYR = (LDH_Vmax*PYR*NADH)/(PYR_Km*NADH_Km+NADH_Km*PYR+PYR_Km*NADH);
19  end

21 % Euler equation used to calculate the concentration of ATP, ADP,
    Pyruvate, and NADH
% for time l with step size n
23 for i = 1:(l*n)
% Change of ATP concentration is equal to - rate of AK + rate of PK
25  ATP(i+1) = ATP(i) - 1/n*AK(ATP(i),AMP(i),AK_Vmax,ATP_Km,AMP_Km) +
        1/n*PK(ADP(i),PEP(i),PK_Vmax,PK_Km);
% Change of AMP concentration is equal to - rate of AK
27  AMP(i+1) = AMP(i) - 1/n*AK(ATP(i),AMP(i),AK_Vmax,ATP_Km,AMP_Km);
% Change of ADP concentration is equal to + 2x rate of AK - rate of PK
29  ADP(i+1) = ADP(i) + 2*1/n*AK(ATP(i),AMP(i),AK_Vmax,ATP_Km,AMP_Km) -
        1/n*PK(ADP(i),PEP(i),PK_Vmax,PK_Km);
% Change of PEP concentration is equal to - rate of PK
31  PEP(i+1) = PEP(i) - 1/n*PK(ADP(i),PEP(i),PK_Vmax,PK_Km);
% Change of Pyruvate concentration is equal to + rate of PK - rate of
    LDH
33  PYR(i+1) = PYR(i) + 1/n*PK(ADP(i),PEP(i),PK_Vmax,PK_Km) - 1/n*LDH(
        PYR(i),NADH(i),LDH_Vmax,PYR_Km,NADH_Km);
% Change of NADH concentration is equal to - rate of LDH
35  NADH(i+1) = NADH(i) - 1/n*LDH(PYR(i),NADH(i),LDH_Vmax,PYR_Km,
```

```
        NADH_Km);
% time variable for x axis
    time(i+1) = time(i)+1/n;
  end
end
function [] = ALLconc_2(ATP,AMP,ADP,PEP,PYR,NADH,AK_Vmax,ATP_Km,AMP_Km,
    PK_Vmax,PK_Km,LDH_Vmax,PYR_Km,NADH_Km,n,l)

% Set time vector for graph
time = (1:(l*n+1))/l;

% Calculate concentration for all 6 substrates in the coupled assay
[ATPrate AMPrate ADPrate PEPrate PYRrate NADHrate] =
    CoupAssayBi_PKbisubstrate(ATP,AMP,ADP,PEP,PYR,NADH,AK_Vmax,ATP_Km,
    AMP_Km,PK_Vmax,PK_Km,LDH_Vmax,PYR_Km,NADH_Km,l,n);

% Plot of all 6 substrates
plot(time,ATPrate,time,AMPrate,time,ADPrate,time,PEPrate,time,PYRrate,
    time,NADHrate)
axis([0 n 0 1000])
legend('ATP','AMP','ADP','PEP','PYR','NADH')
title('Substrates for Coupled Assay')
xlabel('Time (min)')
ylabel('[C] (uM)')
end

function [] = NADHconc_2(ATP,AMP,ADP,PEP,PYR,NADH,AK_Vmax,ATP_Km,AMP_Km,
    PK_Vmax,PK_Km,LDH_Vmax,PYR_Km,NADH_Km,n,l)

% Set number of initial ATP concentrations for calcuations
```

```
ll = size(ATP);
% Set time vector for graph
time = (1:(l*n+1))/l;

% Loop for all initial ATP concentrations
for i = 1:ll(1,2)
% Calculate ATP and NADH concentrations for each initial ATP
    concentrations
    [ATPrate_i , ~ , ~, ~ , ~, NADHrate_i] = CoupAssayBi_PKbisubstrate(
        ATP(1,i),AMP,ADP,PEP,PYR,NADH,AK_Vmax,ATP_Km,AMP_Km,PK_Vmax,
        PK_Km,LDH_Vmax,PYR_Km,NADH_Km,l,n);
% Concatenate each vector for [NADH] and [ATP] into one matrix for each
    ATPrate(i,:) = ATPrate_i;
    NADHrate(i,:) = NADHrate_i;
end

% Plots of NADH concentration at each initial ATP concentrations
figure
plot(time,NADHrate)
axis([0 n 0 max(NADH)])
title('[NADH] rate vs [ATP]')
legend('1000 uM','750 uM','500 uM','100 uM','50 uM','10 uM','5 uM')
xlabel('Time (min)')
ylabel('[NADH] (uM)')

% Plots of ATP concentration at each initial ATP concentrations
figure
plot(time,ATPrate)
axis([0 n 0 max(ATP)])
title('[ATP] rate vs [ATP]')
```

```
87 legend('1000 uM','750 uM','500 uM','100 uM','50 uM','10 uM','5 uM')
   xlabel('Time (min)')
89 ylabel('[ATP] (uM)')

91 end


93

   function[] = CoupledAssay_2()

95

   % Set the time variable (n) in minutes and the number of cycles (l)
97 n = 2;
   l = 1000;
99

   % Set initial concentrations as well as Km and Vmax variables
101 ATP_0   = [1000 750 500 100 50 10 5];
   AMP_0   = 1400;
103 ADP_0   = 0;
   PEP_0   = 2000;
105 PYR_0   = 0;
   NADH_0  = 300;
107 AK_Vmax = 16.03*5;
   ATP_Km  = 0.92;
109 AMP_Km  = 0.92;
   PK_Vmax = 0.68*24000;
111 PK_Km   = 3026;
   LDH_Vmax= 1.71*500;
113 PYR_Km  = 142.7;
   NADH_Km = 142.7;
115
```

```
% Plot of all 6 substrates at 1 initial ATP concentration
figure
ALLconc_2(ATP_0(1),AMP_0,ADP_0,PEP_0,PYR_0,NADH_0,AK_Vmax,ATP_Km,AMP_Km,
    PK_Vmax,PK_Km,LDH_Vmax,PYR_Km,NADH_Km,n,l);

% Plots of ATP and NADH concentrations at all initial ATP concentrations
NADHconc_2(ATP_0,AMP_0,ADP_0,PEP_0,PYR_0,NADH_0,AK_Vmax,ATP_Km,AMP_Km,
    PK_Vmax,PK_Km,LDH_Vmax,PYR_Km,NADH_Km,n,l);
```

Listing B.1: Coupled Assay MATLAB code

B.2 R scripts for Growth Rate Functions

The following R script was used to fit the growth rate function parameters A, B, and D from growth rate and enzyme kinetic measurements of the original 7 mutants and WT TetX2, and produce predictions of $K_{M(MCN)}$ and $V_{max}/K_{M(MCN)}$ of the 9 TetX2 variants we have identified. A second R script was used to estimate variability of the growth rate function predictions using the Jackknife method, and is presented in appendix B.3.

```
rm(list=ls())
#Read data for growth rates and enzyme kinetics.
grates = read.table('GrowthRates_Norm3-2.txt',header = TRUE)
kinetics  = read.table('TetX2_kinetics3.txt',header = TRUE)
run = 'B'

#Set up variables and constants.
bgr = grates[,10]
min = grates[,1]
#Kinetics constants.
```

```
kcat = kinetics[,3]
tetx = kinetics[,5]/kinetics[,3]
knadph = kinetics[,2]
kmin = kinetics[,1]
grmax = 1.0
#Growth rate data.
x1 = grates[,1]
y2 = grates[,2:10]
#Arcsine transform the growth rate data.
y1 = asin((grates[,2:10])^0.5)
y1[is.nan(y1)] = asin(1)
#Set global constants for NADPH concentration and the predicted
    accumulation of MCN inside the cell
nadph = 150
alpha = 4

#Definition of the growth rate curve using a Hill equation; function is
    the growth rate dependent on MCN inside the cell, calculated as
    either the arcsin transformed or the untrasnformed data.
MinGrowth = function(pargr,min){asin((1-1/(1+pargr[1]/min^pargr[2]))
    ^0.5)}
MinGrowth_asin2 = function(pargr,min){1-1/(1+pargr[1]/min^pargr[2])}

#Set up equations for calculating MCN concentration inside the cell.
#Equations for KM(obs), and Vmax(obs) using either the product of kcat
    and tetx concentration or a single reaction velocity.
TetX_Vmax = function(D,kcat,tetx,knadph,nadph){(kcat*tetx/D)/(1+knadph/
    nadph)}
TetX_Vmax_2 = function(D,kcat,knadph,nadph){(kcat/D)/(1+knadph/nadph)}
TetX_kM = function(kmin,knadph,nadph){kmin/(1+knadph/nadph)}
```

```
#Quadratic equation for cytosolic MCN concentration for either a vector
    or scalar calculation.
MIN_Free = function(a,b,c){(-b-(b^2-4*a*c)^0.5)/(2*a)}
MIN_Free2 = function(vmax,km,min0,alpha){(-(min0-km/alpha-vmax)-((min0-
    km/alpha-vmax)^2+4/alpha*km*min0)^0.5)/(-2/alpha)}

#Set up functions for calculating cytosolic minocycline concentration.
#function using v = tetx times kcat.
fmin <- function(p,kcat,tetx,nadph,knadph,kmin,min0){
  vmax <- TetX_Vmax(p,kcat,tetx,knadph,nadph)
  km <- TetX_kM(kmin,knadph,nadph)
  km2 <- km/alpha
  a <- -1/alpha
  b <- outer(km2+vmax,min0,function(x,y){y-x})
  c <- outer(km,min0)
  MIN_Free(a,b,c)
}
#Sum of squares of the difference between measured and predicted growth
    rates.
fsum <- function(p){
  sum((t(y1[,-c]) - MinGrowth(p[1:2],fmin(p[3],kcat[-c],tetx[-c],nadph,
      knadph[-c],kmin[-c],x1)))^2)
}
#Function using only v = kcat
#The two equations are used for to calculate either the matrix or vector
    of growth rate values.
fmin2 <- function(d,kcat_f,nadph_f,knadph_f,kmin_f,min0){
  vmax <- TetX_Vmax_2(d,kcat_f,knadph_f,nadph_f)
  km <- TetX_kM(kmin_f,knadph_f,nadph_f)
  km2 <- km/alpha
```

```
60    a <- -1/alpha
      b <- min0-km2-vmax
62      c <- km*min0
        MIN_Free(a,b,c)
64  }
    fsum2 <- function(p){
66    kmin2 <- p[1]
        kcat2 <- p[1]*p[2]
68      gr2 <- y1[,i]
        sum((t(gr2) - MinGrowth(param[1:2],fmin2(param[3],kcat2,nadph,knadph
            [i],kmin2,x1)))^2)
70  }
    fmin3 <- function(d,kcat_f,nadph_f,knadph_f,kmin_f,min0){
72    vmax <- TetX_Vmax_2(d,kcat_f,knadph_f,nadph_f)
      km <- TetX_kM(kmin_f,knadph_f,nadph_f)
74    km2 <- km/alpha
      a <- -1/alpha
76    b <- min0-km2-vmax
        c <- km*min0
78      MIN_Free(a,b,c)
    }
80  fsum3 <- function(p){
        kmin2 <- p
82      kcat2 <- kact_m[i]*p
        gr2 <- y1[,i]
84      sum((t(gr2) - MinGrowth(param[1:2],fmin3(param[3],kcat2,nadph,knadph
            [i],kmin2,x1)))^2)
    }
86  fmin4 <- function(p,kcat,nadph,knadph,kmin,min0){
      vmax <- TetX_Vmax_2(p,kcat,knadph,nadph)
```

```
  km <- TetX_kM(kmin,knadph,nadph)
  km2 <- km/alpha
  a <- -1/alpha
  b <- outer(km2+vmax,min0,function(x,y){y-x})
  c <- outer(km,min0)
  MIN_Free(a,b,c)
}
fsum4 <- function(p){
  sum((t(y1) - MinGrowth(p[1:2],fmin4(p[3],kcat,nadph,knadph,kmin,x1)))
      ^2)
}

#Model fitting using non-linear least squares of the A, B, and D
    parameters, removing mutant 9, T280D.
param = NULL
c = 9
out = optim(fsum,p = c(20,6,0.01),hessian = 1);
print(out$convergence)
param = out$par
print(param[1:2])
print(param[3])
p <- param

#Model fitting for the kinetic parameters of all 9 mutants.
#Initial fit for KM MCN and the ratio vmax/KM MCN using a set of 2500
    initial conditions.
kmin_p <- NULL
kact_p <- NULL
index <- NULL
minsum <- NULL
```

```
n = 50
for (k in 1:n){
  for (j in 1:n){
            kmin_pt <- NULL
            kact_pt <- NULL
            minsum_t <- NULL
            for (i in 1:length(kcat)){
              out2 = optim(fsum2,p = c((k)*2+10,(j)*0.001),hessian = 1);
              #print(c(out2$convergence,k,j))
              kmin_pt <- rbind(kmin_pt,out2$par[1])
              kact_pt <- rbind(kact_pt,out2$par[2])
              minsum_t <- rbind(minsum_t,out2$value)
            }
            index <- rbind(index,c((k)*2+10,(j)*0.001))
            kmin_p <- cbind(kmin_p,kmin_pt)
            kact_p <- cbind(kact_p,kact_pt)
            minsum <- cbind(minsum,minsum_t)
       }
}

#Find the solution that best fits the maximum slope of the original
    curve.
#First determinie the slope of the meeasured growth rate curve.
varB <- NULL
for (i in 1:(length(x1)-4)){
  varb <- (colSums(x1[i:(i+4)]*y1[i:(i+4),]) - 1/5*(colSums(y1[i:(i+4)
      ,])*sum(x1[i:(i+4)])))/(sum(x1[i:(i+4)]^2) - 1/5*(sum(x1[i:(i+4)])
      )^2)
       varB <- rbind(varB,varb)
}
```

```
varA <- NULL
varC <- NULL
#Determinne the slopes of the 2500 individual solutions for each mutant.
for (l in 1:length(kmin)){
  xp <- x1[which.min(varB[,l]):(which.min(varB[,l])+4)]
  #xp <- x1[which.min((y2[1:8,l]-0.90)^2):(which.min((y2[1:8,l]-0.90)^2)
      +3)]
  varC[l] <- min(varB[,l])
  #varC[l] <- varB[which.min((y2[1:8,l]-0.90)^2),l]
      yp <- t(MinGrowth(p[1:2],fmin4(p[3],kact_p[l,]*kmin_p[l,],nadph,
          knadph[l],kmin_p[l,],xp)))
      vara <- (colSums(xp*yp) - 1/5*(colSums(yp)*sum(xp)))/(sum(xp^2) -
          1/5*(sum(xp)^2))
      varA <- cbind(varA,vara)
}
#Find the solution that best fits the slope to determine the predicted
    vmax/KM MCN.
colnames(varA) <- colnames(y1)
kmin_ps <- NULL
kact_ps <- NULL
fitbeta <- NULL
for(i in 1:length(kcat)){
      fitbeta[i] <- which.min((varA[,i] - varC[i])^2)
      kmin_ps <- rbind(kmin_ps,kmin_p[i,fitbeta[i]])
      kact_ps <- rbind(kact_ps,kact_p[i,fitbeta[i]])
}

#Print the original parameter fit solution curves and the set of
    solution curves from the first two steps for the kinetic parameters
    fit.
```

```
166 tiff(file = paste(Sys.Date(),run,'solutions',sep = '_'),
    width = 16.0, height = 8.7, units = 'cm',
168 res = 800, antialias = 'default')

170 mutnames = names(grates)
par(mfrow = c(2,5), oma = c(4,4,1,1),mar = c(0,0,1,0),xpd = NA)
172 lbl = c('y','n','n','n','n','b','x','x','x')
for (l in 1:length(kcat)){
174 plot(grates[,1],grates[,l+1], xlim = (c(0,grates[length(x1),1])), ylim
      = c(0,1),
          labels = F, frame.plot = T, xlab = NA, ylab = NA, xaxp = c
              (0,80,4), pch = 20, cex = 0.66);
176 for(i in 1:length(kmin_p[l,])){
    curve(MinGrowth_asin2(p[1:2],fmin3(p[3],kact_p[l,i]*kmin_p[l,i],
        nadph,knadph[l],kmin_p[l,i],x)),
178 0, max(x1), add = T, lwd = 0.1)
  }
180

  curve(MinGrowth_asin2(p[1:2],fmin3(p[3],kact_ps[l]*kmin_ps[l],nadph,
      knadph[l],kmin_ps[l],x)),
182 0, max(x1), add = T, lwd = 0.3, col = 2)
  mtext(paste(mutnames[l+1],'  '), outer = F, side = 3, adj = 1, line
      = -1.5, cex = 0.6)
184 if (lbl[l] == 'x'){
    axis(side = 1, tick = F, cin = 6/72)
186 } else if (lbl[l] == 'y'){
    axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
188 } else if (lbl[l] == 'b'){
    axis(side = 1, tick = F)
190 axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
```

```
    }
}
mtext(paste('A)'), line = 1.2, outer = T, cin = 6/72, adj = 0)
mtext(expression(paste('Minocycline ','(', mu,'g/ml)')), outer = T, side
    = 1, line = 2.2, cin = 6/72)
mtext('Growth Rate', outer = T, side = 2, line = 2.4)
dev.off()

tiff(file = paste(Sys.Date(),run,'ratios',sep = '_'),
  width = 16.0, height = 8.7, units = 'cm',
  res = 800, antialias = 'default')
mutnames = names(grates)
par(mfrow = c(2,5), oma = c(4,4,1,1),mar = c(0,0,1,0),xpd = NA)
lbl = c('y','n','n','n','n','b','x','x','x')
for (l in 1:length(kcat)){
  plot(grates[,1],grates[,l+1], xlim = (c(0,grates[length(x1),1])), ylim
      = c(0,1),
    labels = F, frame.plot = T, xlab = NA, ylab = NA, xaxp = c(0,80,4),
       pch = 20, cex = 0.66, col = 3)
  for(i in 1:1000){
    curve(MinGrowth_asin2(p[1:2],fmin3(p[3],kact_ps[l]*kmin_ps[l]*(i/
       100),nadph,knadph[l],kmin_ps[l]*(i/100),x)),
          0, max(x1), add = T, lwd = 0.1)
  }
  curve(MinGrowth_asin2(p[1:2],fmin3(p[3],kact_ps[l]*kmin_ps[l],nadph,
      knadph[l],kmin_ps[l],x)),
       0, max(x1), add = T, lwd = 0.3, col = 2)
  mtext(paste(mutnames[l+1],'  '), outer = F, side = 3, adj = 1.2, line
      = -1.2, cex = 0.6)
  if (lbl[l] == 'x'){
```

```
    axis(side = 1, tick = F, cin = 6/72)
  } else if (lbl[l] == 'y'){
    axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
  } else if (lbl[l] == 'b'){
    axis(side = 1, tick = F)
    axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
  }
}
mtext(paste('A)'), line = 1.2, outer = T, cin = 6/72, adj = 0)
mtext(expression(paste('Minocycline ','(', mu,'g/ml)')), outer = T, side
    = 1, line = 2.2, cin = 6/72)
mtext('Growth Rate', outer = T, side = 2, line = 2.4)
dev.off()

#Fit for KM MCN using a second least squares fit with the ratio of vmax/
   KM MCN determined from the previous fit.
kact_m <- kact_ps
#kact_m <- apply(kact_p,1,mean)

kmin_pd <- NULL
kcat_pd <- NULL
for(i in 1:9){
  out3 = optim(fsum3,p = 50,hessian = 1);
  print(c(out3$convergence))
  kmin_pd <- rbind(kmin_pd,out3$par[1])
  kcat_pd <- rbind(kcat_pd,out3$par[1]*kact_m[i])
}

#Print curves of the predicted growth rates using the predicted kinetic
   parameters for each mutant.
```

```
tiff(file = paste(Sys.Date(),run,'solution-final',sep = '_'),
  width = 16.0, height = 8.7, units = 'cm',
  res = 800, antialias = 'default')

mutnames = names(grates)
par(mfrow = c(2,5), oma = c(4,4,1,1),mar = c(0,0,1,0),xpd = NA)
lbl = c('y','n','n','n','n','b','x','x','x')
for (l in 1:length(kcat)){
  plot(grates[,1],grates[,l+1], xlim = (c(0,grates[length(x1),1])), ylim
      = c(0,1),
    labels = F, frame.plot = T, xlab = NA, ylab = NA, xaxp = c(0,80,4),
      pch = 20, cex = 0.66);
  curve(MinGrowth_asin2(p[1:2],fmin3(p[3],kcat_pd[l],nadph,knadph[l],
    kmin_pd[l],x)),
    0, max(x1), add = T, lwd = 0.3, col = 2)
  mtext(paste(mutnames[l+1],'  '), outer = F, side = 3, adj = 1, line =
    -1.5, cex = 0.6)
  if (lbl[l] == 'x'){
  axis(side = 1, tick = F, cin = 6/72)
  } else if (lbl[l] == 'y'){
  axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
  } else if (lbl[l] == 'b'){
  axis(side = 1, tick = F)
  axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
  }
}
mtext(paste('A)'), line = 1.2, outer = T, cin = 6/72, adj = 0)
mtext(expression(paste('Minocycline ','(', mu,'g/ml)')), outer = T, side
    = 1, line = 2.2, cin = 6/72)
mtext('Growth Rate', outer = T, side = 2, line = 2.4)
```

```
dev.off()

#Write and print results
Results <- data.frame('KM MCN' = kmin_pd, 'Vmax' = kcat_pd, 'Activity' =
    kact_m)
rownames(Results) <- mutnames[2:10]

print(Results)

write.table(Results, file = paste(Sys.Date(),run,'Results',sep = '_'),
   sep = ',', row.names = T, col.names = T)

tiff(file = paste(Sys.Date(),run,'Figure5-solutions',sep = '_'),
  width = 5.5, height = 5.5, units = 'cm',
  res = 1200, antialias = 'default')

mutnames = names(grates)
par(mfrow = c(2,5), oma = c(4,4,1,1),mar = c(0,0,1,0),xpd = NA)
lbl = c('y','n','n','n','n','b','x','x','x')
for (l in 1:length(kcat)){
  plot(grates[,1],grates[,l+1], xlim = (c(0,grates[length(x1),1])), ylim
      = c(0,1),
          labels = F, frame.plot = T, xlab = NA, ylab = NA, xaxp = c
              (0,80,4), pch = 1, cex = 0.66);
        curve(MinGrowth_asin2(p[1:2],fmin(p[3],kcat[l],tetx[l],nadph,
            knadph[l],kmin[l],x)),
          0, max(x1), add = T, lwd = 1)
  curve(MinGrowth_asin2(p[1:2],fmin3(p[3],kact_ps[l]*kmin_ps[l],nadph,
      knadph[l],kmin_ps[l],x)),
          0, max(x1), add = T, lwd = 1, lty = 'dashed', col = 2)
```

```
    mtext(paste(mutnames[l+1],'   '), outer = F, side = 3, adj = 1.2,
        line = -0.8, cex = 0.4)
    if (lbl[l] == 'x'){
      axis(side = 1, tick = F, cin = 6/72)
    } else if (lbl[l] == 'y'){
      axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
    } else if (lbl[l] == 'b'){
      axis(side = 1, tick = F)
      axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
    }
}
mtext(paste('A)'), line = 1.2, outer = T, cin = 6/72, adj = 0)
mtext(expression(paste('Minocycline ','(', mu,'g/ml)')), outer = T, side
    = 1, line = 2.2, cin = 6/72)
mtext('Growth Rate', outer = T, side = 2, line = 2.4)
dev.off()
```

Listing B.2: R script for Growth Rate Function of TetX2

B.3 R Script for Jackknife Method

```
rm(list=ls())
#Read data for growth rates and kinetics.
grates = read.table('GrowthRates_Norm3-2.txt',header = TRUE)
kinetics  = read.table('TetX2_kinetics3-2.txt',header = TRUE)
runs = c('B1','B2','B3','B4','B5','B6','B7','B8')

#Set up variables and constants.
bgr = grates[,10]
min = grates[,1]
```

```
#Kinetics constants.
kcat = kinetics[,3]
tetx = kinetics[,5]/kinetics[,3]
knadph = kinetics[,2]
kmin = kinetics[,1]
grmax = 1.0
#Growth rate data.
x1 = grates[,1]
y2 = grates[,2:9]
y1 = asin((grates[,2:9])^0.5)
y1[is.nan(y1)] = asin(1)
nadph = 150
alpha = 4

#Fit the no-TetX growth curve using a hill equation; function is the
    growth rate dependent on minocycline inside the cell.
MinGrowth = function(pargr,min){asin((1-min^pargr[2]/(pargr[1]+min^pargr
    [2]))^0.5)}
MinGrowth_asin2 = function(pargr,min){1-min^pargr[2]/(pargr[1]+min^pargr
    [2])}

#Set up equations for calculating minocycline concentration inside the
    cell.
TetX_Vmax = function(D,kcat,tetx,knadph,nadph){(kcat*tetx/D)*nadph/(
    knadph+nadph)}
TetX_Vmax_2 = function(D,kcat,knadph,nadph){(kcat/D)*nadph/(knadph+nadph
    )}
TetX_kM = function(kmin,knadph,nadph){kmin*nadph/(knadph+nadph)}
MIN_Free = function(a,b,c){(-b-(b^2-4*a*c)^0.5)/(2*a)}
```

```
MIN_Free2 = function(vmax,km,min0,alpha){(-(min0-km/alpha-vmax)-((min0-
    km/alpha-vmax)^2+4/alpha*km*min0)^0.5)/(-2/alpha)}

#Fit all data with dr, nadph, and tetx as variables. tetx can vary
    between mutants.
fmin <- function(p,kcat,tetx,nadph,knadph,kmin,min0){
  vmax <- TetX_Vmax(p,kcat,tetx,knadph,nadph)
  km <- TetX_kM(kmin,knadph,nadph)
  km2 <- km/alpha
  a <- -1/alpha
  b <- outer(km2+vmax,min0,function(x,y){y-x})
  c <- outer(km,min0)
  MIN_Free(a,b,c)
}
fsum <- function(p){
  sum((t(y1[,-c]) - MinGrowth(p[1:2],fmin(p[3],kcat[-c],tetx[-c],nadph,
      knadph[-c],kmin[-c],x1)))^2)
}
fmin2 <- function(d,kcat_f,nadph_f,knadph_f,kmin_f,min0){
  vmax <- TetX_Vmax_2(d,kcat_f,knadph_f,nadph_f)
  km <- TetX_kM(kmin_f,knadph_f,nadph_f)
  km2 <- km/alpha
  a <- -1/alpha
  b <- min0-km2-vmax
    c <- km*min0
    MIN_Free(a,b,c)
}
fsum2 <- function(p){
  kmin2 <- p[1]
    kcat2 <- p[1]*p[2]
```

```
60      gr2 <- y1[,i]
        sum((t(gr2) - MinGrowth(param[1:2],fmin2(param[3],kcat2,nadph,knadph
            [i],kmin2,x1)))^2)
62  }
    fmin3 <- function(d,kcat_f,nadph_f,knadph_f,kmin_f,min0){
64    vmax <- TetX_Vmax_2(d,kcat_f,knadph_f,nadph_f)
      km <- TetX_kM(kmin_f,knadph_f,nadph_f)
66    km2 <- km/alpha
      a <- -1/alpha
68    b <- min0-km2-vmax
        c <- km*min0
70      MIN_Free(a,b,c)
    }
72  fsum3 <- function(p){
        kmin2 <- p
74      kcat2 <- kact_m[i]*p
        gr2 <- y1[,i]
76      sum((t(gr2) - MinGrowth(param[1:2],fmin3(param[3],kcat2,nadph,knadph
            [i],kmin2,x1)))^2)
    }
78  fmin4 <- function(p,kcat,nadph,knadph,kmin,min0){
      vmax <- TetX_Vmax_2(p,kcat,knadph,nadph)
80    km <- TetX_kM(kmin,knadph,nadph)
      km2 <- km/alpha
82    a <- -1/alpha
      b <- outer(km2+vmax,min0,function(x,y){y-x})
84    c <- outer(km,min0)
      MIN_Free(a,b,c)
86  }
    fsum4 <- function(p){
```

```
  sum((t(y1) - MinGrowth(p[1:2],fmin4(p[3],kcat,nadph,knadph,kmin,x1)))
      ^2)
}

parameters <- NULL
for(c in c(1:8)){
  #Model fitting using non-linear least squares.
  param = NULL
  out = optim(fsum,p = c(20,2.3,0.001), hessian = 1);
  print(out$convergence)
  param = out$par
  print(param[1:2])
  print(param[3])
  p <- param
  parameters <- rbind(parameters,p)
  run <- runs[c]
  assign(paste('out', run, sep = '_'),out)


  kmin_p <- NULL
  kact_p <- NULL
  index <- NULL
  minsum <- NULL
  n = 50
  for (k in 1:n){
    for (j in 1:n){
            kmin_pt <- NULL
            kact_pt <- NULL
            minsum_t <- NULL
            for (i in 1:length(kcat)){
```

```
                out2 = optim(fsum2,p = c((k)*2+10,(j)*0.001+0.001),hessian
                    = 1);
                print(c(out2$convergence,k,j))
                kmin_pt <- rbind(kmin_pt,out2$par[1])
                kact_pt <- rbind(kact_pt,out2$par[2])
                minsum_t <- rbind(minsum_t,out2$value)
              }
              index <- rbind(index,c((k)*2+10,(j)*0.001))
              kmin_p <- cbind(kmin_p,kmin_pt)
              kact_p <- cbind(kact_p,kact_pt)
              minsum <- cbind(minsum,minsum_t)
        }
}

varB <- NULL
for (i in 1:(length(x1)-4)){
  varb <- (colSums(x1[i:(i+4)]*y1[i:(i+4),]) - 1/5*(colSums(y1[i:(i+4)
      ,])*sum(x1[i:(i+4)])))/(sum(x1[i:(i+4)]^2) - 1/5*(sum(x1[i:(i+4)
      ]))^2)
       varB <- rbind(varB,varb)
}
varA <- NULL
varC <- NULL
y3 <- NULL
for (i in 2:8){
  yt <- colMeans(y2[i-1:i+1,])
  y3 <- rbind(y3,yt)
}
for (l in 1:length(kmin)){
  xp <- x1[which.min(varB[,l]):(which.min(varB[,l])+4)]
```

```
    #xp <- x1[(which.min((y2[3:6,l]-0.5)^2)):(which.min((y2[3:6,l]-0.5)^2)+4)]
    varC[l] <- min(varB[,l])
    #varC[l] <- varB[which.min((y2[3:6,l]-0.5)^2)+2,l]
        yp <- t(MinGrowth(p[1:2],fmin4(p[3],kact_p[l,]*kmin_p[l,],nadph,knadph[l],kmin_p[l,],xp)))
        vara <- (colSums(xp*yp) - 1/5*(colSums(yp)*sum(xp)))/(sum(xp^2) - 1/5*(sum(xp)^2))
        varA <- cbind(varA,vara)
}
colnames(varA) <- colnames(y1)
kmin_ps <- NULL
kact_ps <- NULL
fitbeta <- NULL
for(i in 1:length(kcat)){
        fitbeta[i] <- which.min((varA[,i] - varC[i])^2)
        kmin_ps <- rbind(kmin_ps,kmin_p[i,fitbeta[i]])
        kact_ps <- rbind(kact_ps,kact_p[i,fitbeta[i]])
}

tiff(file = paste(Sys.Date(),run,'solutions',sep = '_'),
  width = 16.0, height = 8.7, units = 'cm',
  res = 800, antialias = 'default')

mutnames = names(grates)
par(mfrow = c(2,4), oma = c(4,4,1,1),mar = c(0,0,1,0),xpd = NA)
lbl = c('y','n','n','n','b','x','x','x')
for (l in 1:length(kcat)){
  plot(grates[,1],grates[,l+1], xlim = (c(0,grates[length(x1),1])),
      ylim = c(0,1),
```

```
            labels = F, frame.plot = T, xlab = NA, ylab = NA, xaxp = c
                (0,80,4), pch = 20, cex = 0.66);
        for(i in 1:length(kmin_p[l,])){
          curve(MinGrowth_asin2(p[1:2],fmin3(p[3],kact_p[l,i]*kmin_p[l,i],
              nadph,knadph[l],kmin_p[l,i],x)),
                0, max(x1), add = T, lwd = 0.1)
        }

        curve(MinGrowth_asin2(p[1:2],fmin3(p[3],kact_ps[l]*kmin_ps[l],
            nadph,knadph[l],kmin_ps[l],x)),
                0, max(x1), add = T, lwd = 0.3, col = 2)
        mtext(paste(mutnames[l+1],'  '), outer = F, side = 3, adj = 1,
            line = -1.5, cex = 0.6)
        if (lbl[l] == 'x'){
          axis(side = 1, tick = F, cin = 6/72)
        } else if (lbl[l] == 'y'){
          axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
        } else if (lbl[l] == 'b'){
          axis(side = 1, tick = F)
          axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
        }
    }
    mtext(paste('A)'), line = 1.2, outer = T, cin = 6/72, adj = 0)
    mtext('Minocycline (ug/ml)', outer = T, side = 1, line = 2.2, cin = 6/
        72)
    mtext('Growth Rate', outer = T, side = 2, line = 2.4)
    dev.off()

    tiff(file = paste(Sys.Date(),run,'ratios',sep = '_'),
      width = 16.0, height = 8.7, units = 'cm',
```

```
    res = 800, antialias = 'default')
196 mutnames = names(grates)
par(mfrow = c(2,4), oma = c(4,4,1,1),mar = c(0,0,1,0),xpd = NA)
198 lbl = c('y','n','n','n','b','x','x','x')
for (l in 1:length(kcat)){
200   plot(grates[,1],grates[,l+1], xlim = (c(0,grates[length(x1),1])),
        ylim = c(0,1),
      labels = F, frame.plot = T, xlab = NA, ylab = NA, xaxp = c(0,80,4)
          , pch = 20, cex = 0.66, col = 3)
202   for(i in 1:1000){
      curve(MinGrowth_asin2(p[1:2],fmin3(p[3],kact_ps[l]*kmin_ps[l]*(i/
          100),nadph,knadph[l],kmin_ps[l]*(i/100),x)),
204           0, max(x1), add = T, lwd = 0.1)
    }
206   curve(MinGrowth_asin2(p[1:2],fmin3(p[3],kact_ps[l]*kmin_ps[l],nadph,
        knadph[l],kmin_ps[l],x)),
        0, max(x1), add = T, lwd = 0.3, col = 2)
208   mtext(paste(mutnames[l+1],'  '), outer = F, side = 3, adj = 1, line
        = -1.5, cex = 0.6)
    if (lbl[l] == 'x'){
210   axis(side = 1, tick = F, cin = 6/72)
    } else if (lbl[l] == 'y'){
212   axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
    } else if (lbl[l] == 'b'){
214   axis(side = 1, tick = F)
      axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
216   }
}
218 mtext(paste('A)'), line = 1.2, outer = T, cin = 6/72, adj = 0)
```

```
mtext('Minocycline (ug/ml)', outer = T, side = 1, line = 2.2, cin = 6/
    72)
mtext('Growth Rate', outer = T, side = 2, line = 2.4)
dev.off()

kact_m <- kact_ps
#kact_m <- apply(kact_p,1,mean)

kmin_pd <- NULL
kcat_pd <- NULL
for(i in 1:8){
  out3 = optim(fsum3,p = 50,hessian = 1);
  print(c(out3$convergence))
  kmin_pd <- rbind(kmin_pd,out3$par[1])
  kcat_pd <- rbind(kcat_pd,out3$par[1]*kact_m[i])
}

tiff(file = paste(Sys.Date(),run,'solution-final',sep = '_'),
  width = 16.0, height = 8.7, units = 'cm',
  res = 800, antialias = 'default')

mutnames = names(grates)
par(mfrow = c(2,4), oma = c(4,4,1,1),mar = c(0,0,1,0),xpd = NA)
lbl = c('y','n','n','n','b','x','x','x')
for (l in 1:length(kcat)){
  plot(grates[,1],grates[,l+1], xlim = (c(0,grates[length(x1),1])),
      ylim = c(0,1),
    labels = F, frame.plot = T, xlab = NA, ylab = NA, xaxp = c(0,80,4)
        , pch = 20, cex = 0.66);
```

```
    curve(MinGrowth_asin2(p[1:2],fmin3(p[3],kcat_pd[l],nadph,knadph[l],
        kmin_pd[l],x)),
246   0, max(x1), add = T, lwd = 0.3, col = 2)
    mtext(paste(mutnames[l+1],'  '), outer = F, side = 3, adj = 1, line
        = -1.5, cex = 0.6)
248   if (lbl[l] == 'x'){
    axis(side = 1, tick = F, cin = 6/72)
250   } else if (lbl[l] == 'y'){
    axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
252   } else if (lbl[l] == 'b'){
    axis(side = 1, tick = F)
254   axis(side = 2, tick = F, at = c(0,0.5,1), cin = 6/72)
    }
256 }
  mtext(paste('A)'), line = 1.2, outer = T, cin = 6/72, adj = 0)
258 mtext('Minocycline (ug/ml)', outer = T, side = 1, line = 2.2, cin = 6/
      72)
  mtext('Growth Rate', outer = T, side = 2, line = 2.4)
260 dev.off()

262 Results <- data.frame('KM MCN' = kmin_pd, 'Vmax' = kcat_pd, 'Activity'
      = kact_m)
  rownames(Results) <- mutnames[2:9]
264 assign(paste('Results', run, sep = '_'),Results)

266 write.table(Results, file = paste(Sys.Date(),run,'Results',sep = '_'),
      sep = ',', row.names = T, col.names = T)
}
268

print(Results_B1)
```

```
print(Results_B2)
print(Results_B3)
print(Results_B4)
print(Results_B5)
print(Results_B6)
print(Results_B7)
print(Results_B8)
print(parameters)

write.table(parameters, file = paste(Sys.Date(),'B','Parameters',sep = '
    _'), sep = ',', row.names = F, col.names = F)
```

Listing B.3: R script for Growth Rate Function of TetX2

Appendix C

Jackknife Results

Table C.1: Estimate of kinetic parameters using jackknife method model parameters

TetX2 removed	$K_{M(MCN)}$ (μM)	V_{max} (s^{-1})	$V_{max}/K_{M(MCN)}$ ($\mu M^{-1}\ s^{-1}$)
TetX2	28.62	0.1	0.0035
T280A	5.93	0.12	0.02
F235Y	101.45	0.56	0.0055
K64R	13.02	0.06	0.0046
N371I	29.61	0.4	0.0136
N371T	19.11	0.25	0.0133
S326I	96.56	0.6	0.0062

T280S	18.79	0.07	0.0039
T280A removed	$K_{M(MCN)}$ (μM)	V_{max} (s^{-1})	$V_{max}/K_{M(MCN)}$ ($\mu M^{-1} s^{-1}$)
TetX2	23.68	0.09	0.004
T280A	5.09	0.12	0.0229
F235Y	74.02	0.47	0.0063
K64R	20.37	0.08	0.004
N371I	32.11	0.46	0.0144
N371T	22.14	0.29	0.0132
S326I	83.2	0.58	0.0069
T280S	18.96	0.08	0.0041
F235Y removed	$K_{M(MCN)}$ (μM)	V_{max} (s^{-1})	$V_{max}/K_{M(MCN)}$ ($\mu M^{-1} s^{-1}$)
TetX2	26.74	0.11	0.0041
T280A	1.24	0.02	0.0194
F235Y	1966125	14659.72	0.0075

K64R	62.03	0.22	0.0036
N371I	3932205	56785.22	0.0144
N371T	5242925	56382.73	0.0108
S326I	3932205	31277.96	0.008
T280S	18.34	0.07	0.0039

K64R removed	$K_{M(MCN)}$ (μM)	V_{max} (s^{-1})	$V_{max}/K_{M(MCN)}$ ($\mu M^{-1}\ s^{-1}$)
TetX2	15.8	0.07	0.0045
T280A	5.75	0.11	0.0199
F235Y	87.91	0.51	0.0058
K64R	13.13	0.06	0.0046
N371I	41.38	0.53	0.0128
N371T	20.57	0.26	0.0127
S326I	576.25	3.37	0.0058
T280S	18.54	0.07	0.004

N371I removed	$K_{M(MCN)}$	V_{max}	$V_{max}/K_{M(MCN)}$

	(μM)	(s^{-1})	($\mu M^{-1} s^{-1}$)
TetX2	18	0.08	0.0044
T280A	5.92	0.12	0.0201
F235Y	84.65	0.52	0.0061
K64R	19.45	0.08	0.004
N371I	60.94	0.77	0.0126
N371T	16.21	0.24	0.0146
S326I	101.52	0.68	0.0067
T280S	16.8	0.07	0.0043

N371T removed	$K_{M(MCN)}$ (μM)	V_{max} (s^{-1})	$V_{max}/K_{M(MCN)}$ ($\mu M^{-1} s^{-1}$)
TetX2	32.99	0.12	0.0036
T280A	6.52	0.13	0.0203
F235Y	111.91	0.65	0.0058
K64R	19.45	0.08	0.0041
N371I	36.07	0.5	0.0138

N371T	21.68	0.29	0.0134
S326I	93.13	0.62	0.0067
T280S	14.82	0.07	0.0047

S236I removed	$K_{M(MCN)}$ (μM)	V_{max} (s^{-1})	$V_{max}/K_{M(MCN)}$ ($\mu M^{-1} s^{-1}$)
TetX2	35.25	0.13	0.0036
T280A	4.68	0.09	0.0195
F235Y	45.65	0.32	0.007
K64R	16.22	0.07	0.0042
N371I	53.34	0.72	0.0134
N371T	19.79	0.25	0.0125
S326I	86.29	0.63	0.0073
T280S	14.87	0.06	0.0043

T280S removed	$K_{M(MCN)}$ (μM)	V_{max} (s^{-1})	$V_{max}/K_{M(MCN)}$ ($\mu M^{-1} s^{-1}$)
TetX2	25.96	0.09	0.0035

T280A	5.31	0.1	0.0187
F235Y	127.23	0.7	0.0055
K64R	13.28	0.06	0.0043
N371I	50.17	0.61	0.0121
N371T	30.87	0.32	0.0104
S326I	64.41	0.43	0.0066
T280S	16.7	0.07	0.0039